

# Využití pravděpodobnostních modelů při analýze CT snímků

Application of probabilistic models in CT image analysis

Matěj Mazůrek

Bakalářská práce

Vedoucí práce: Ing. Jan Kracík, Ph.D.

Ostrava, 2021

## Abstrakt

Lékařské snímky bývají často zaneseny například šumem, který tyto obrázky znehodnocuje. Pravděpodobnostní modelování pak představuje jeden z přístupů k analýze obrazu. V této bakalářské práci se tedy pokusíme sestavit pravděpodobnostní model rozdělení oblastí CT snímku aorty břišní, pomocí kterého se pak pokusíme najít optimální segmentaci takového snímku. Využijeme principů bayesovské statistiky a simulačních algoritmů Markov Chain Monte Carlo, které se pro takové úlohy velice hodí. Pro správné uchopení těchto pojmů a samotné úlohy jsou v práci zavedeny základní pojmy teorie pravděpodobnosti, markovské řetězce, markovská náhodná pole, teorie míry a následně vysvětlení principů Monte Carlo metod.

## Klíčová slova

CT snímek, analýza obrazu, pravděpodobnostní model, markovské náhodné pole, bayesovská statistika, Markov Chain Monte Carlo, Gibbsův sampler

## Abstract

Medical images are often covered with noise, for example, which devalues these images. Probabilistic modelling then represents one of the approaches to image analysis. In this Bachelor thesis we will try to build a model of the distribution of areas of the CT image of the abdominal aorta, with which we will then try to find the optimal segmentation of such an image. We will use the principles of Bayesian statistics and Markov Chain Monte Carlo simulation algorithms, which are well suited for such tasks. To properly grasp these concepts and the task itself, the basic concepts of measure theory, probability theory, Markov chains, Markov random fields, measure theory and then an explanation of the principles of Monte Carlo methods are introduced.

## Klíčová slova

CT image, image analysis, probabilistic model, Markov random field, Bayesian statistics, Markov Chain Monte Carlo, Gibbs sampler

# Obsah

<b>Seznam obrázků</b>	<b>4</b>
<b>1 Úvod</b>	<b>5</b>
<b>2 Teorie pravděpodobnosti</b>	<b>6</b>
2.1 Základní pojmy a definice . . . . .	6
2.2 Nezávislost jevů, podmíněná pravděpodobnost . . . . .	8
2.3 Věta o úplné pravděpodobnosti, Bayesova věta . . . . .	9
<b>3 Náhodná veličina, náhodný vektor</b>	<b>11</b>
3.1 Náhodná veličina a její diskrétní a absolutně spojitě rozdělení . . . . .	11
3.2 Náhodný vektor . . . . .	13
3.3 Číselné charakteristiky náhodné veličiny a náhodného vektoru . . . . .	16
3.4 Normální (Gaussovo) rozdělení náhodné veličiny . . . . .	17
<b>4 Markovský řetězec, markovské náhodné pole</b>	<b>18</b>
4.1 Markovský řetězec . . . . .	18
4.2 Markovská náhodná pole . . . . .	20
4.3 Gibbsovo rozdělení . . . . .	22
<b>5 Markov Chain Monte Carlo</b>	<b>23</b>
5.1 Monte Carlo metody . . . . .	23
5.2 Markov Chain Monte Carlo metody . . . . .	24
<b>6 Teorie míry</b>	<b>26</b>
<b>7 Bayesovská rozhodovací úloha</b>	<b>29</b>
<b>8 Segmentace CT snímku aorty břišní</b>	<b>32</b>
<b>9 Závěr</b>	<b>41</b>
<b>Reference</b>	<b>42</b>

## Seznam obrázků

1	Ukázka úplného systému vzájemně disjunktních jevů a jeho průnik s jevem A . . . .	10
2	Ilustrace hustoty pravděpodobnosti a pravděpodobnostní funkce . . . . .	12
3	Hustota normálního rozdělení v závislosti na parametrech . . . . .	17
4	Znázornění dvou systémů okolí $\mathcal{N}$ neorientovanými grafy . . . . .	21
5	Gibbsův sampler z dvojrozměrného normálního rozdělení pro $N = 10$ a $N = 100$ . .	25
6	Znázornění tvaru aposteriorní hustoty pro rozdílný počet pozorovaných dat . . . . .	31
7	Znázornění segmentací obrázku $6 \times 6$ vlevo s menší energií, vpravo s větší energií . .	33
8	CT snímek aorty břišní . . . . .	36
9	Výsledná segmentace s vhodnými parametry po 15 iteracích . . . . .	37
10	Vlevo segmentace po 1 a vpravo po 5 iteracích . . . . .	37
11	Ukázka stacionárního rozdělení Gibbsova sampleru . . . . .	38
12	Segmentace po 15 iteracích s $\beta = \frac{1}{10}$ . . . . .	38
13	Segmentace se změněnými rozptyly . . . . .	39
14	Segmentace se změněnými středními hodnotami . . . . .	39

# 1 Úvod

Lékařské snímky, převážně CT snímky, bývají často zaneseny šumem či jinými rušivými elementy, což mnohdy znepřístupňuje nebo dokonce znemožňuje učinit důležité medicínské závěry. V návaznosti na probíhající projekt MŠMT Výpočtové statistické hodnocení rizika ruptury aneurysmat abdominální aorty, který se zabývá riziky pacientů s výdutí (aneurysmem) aorty břišní se pokusíme segmentovat anonymizovaný CT snímek aorty, tedy rozdělit jej na tři oblasti - lumen neboli prostor, kterým teče krev, trombus neboli krevní sraženinu uvnitř aorty a na vnější prostor aorty. Tato analýza CT snímku je důležitá právě proto, že výduť cévy představuje životunebezpečnou situaci, kdy céva zduří a je najednou vystavována zvýšenému tlaku krve. V takových případech hrozí například rizikové protrhnutí cévy nebo jiné problémy. Krevní sraženiny v aneurysmech nejsou výjimkou, představují ale další komplikace.

V analýze obrazu je často využíváno pravděpodobnostních modelů založených například na bayesovské statistice a Markov Chain Monte Carlo metodách. Přesně o sestavení takového modelu se pokusíme. Na jeho základě pak pomocí Gibbsova sampleru, speciálního Markov Chain Monte Carlo algoritmu, odhadneme optimální segmentaci CT snímku.

V několika úvodních kapitolách se seznámíme s potřebnými pojmy elementární teorie pravděpodobnosti, což vyústí ve formulaci Bayesovy věty a definicím náhodné veličiny, náhodného vektoru a normálního rozdělení. Dále se seznámíme s pojmy jako markovský řetězec a markovské náhodné pole, které jsou základem pro pochopení a správné zpracování následujících algoritmů a problému. Posléze následuje kapitola o simulačních metodách Monte Carlo a speciálně Markov Chain Monte Carlo, které primárně sice slouží k numerické integraci, ale díky specifickým vlastnostem cílových hustot se jejich vzorkování hodí i pro naši úlohu. Následuje kapitola s úvodem do teorie míry, která upevňuje a svým způsobem kompletuje teorii pravděpodobnosti. Bez aparátu teorie míry bychom jen ztěžili zkonstruovali náš model. Další kapitolou je popis bayesovské rozhodovací úlohy, jejích principů a důležitých pojmů, což tvoří základ pro středobod celé práce, a tím je konečně kapitola segmentace CT snímku aorty břišní. V ní se na základě dostupných informací zvolí vhodná apriorní informace, sestaví se model, určí se aposteriorní hustota a ztrátová funkce a nakonec pomocí Gibbsova sampleru získá optimální segmentace.

## 2 Teorie pravděpodobnosti

### 2.1 Základní pojmy a definice

Pro správné uchopení aparátu pravděpodobnostního modelování je třeba si nejdříve nadefinovat několik důležitých pojmů teorie pravděpodobnosti. Na úvod si zavedme pojmy, které tvoří neformální základ této teorie. V úvodních dvou kapitolách ((2), (3)) bylo čerpáno především ze skript VŠB-TUO [1] a [2].

- **Náhodný pokus** je takový proces, který je (většinou teoreticky) do nekonečna opakovatelný a jehož výsledek není jednoznačně určen okolními podmínkami.
- Uvažujme množinu všech možných výsledků náhodného pokusu  $\Omega$ . **Elementárními jevy** pak nazveme prvky  $\omega \in \Omega$ . Při realizaci náhodného pokusu vždy nastane právě jeden elementární jev.
- **Náhodný jev** je podmnožinou množiny všech možných výsledků náhodného pokusu. Pro náhodný jev  $A$  platí  $A \subset \Omega$ . Po provedení náhodného pokusu můžeme určit, zda jev  $A$  nastal ( $\omega \in A$ ) nebo nenastal ( $\omega \notin A$ ).

Náhodné jevy značíme zpravidla velkými písmeny latinské abecedy. Přiblížme si dále některé vlastnosti a relace náhodných jevů.

Rozlišujeme dva speciální případy náhodných jevů, a to:

- **Jev jistý** je roven  $\Omega$  a je to takový jev, který při uskutečnění náhodného pokusu nastane vždy.
- **Jev nemožný** při realizaci náhodného pokusu nemůže nastat.

Zmíníme také základní operace s náhodnými jevy (při nichž vznikne vždy jev nový), a to:

- **Sjednocení náhodných jevů**  
Sjednocení jevů  $A$  a  $B$  ( $A \cup B$ ) nastává, když nastal jev  $A$  nebo jev  $B$ .
- **Průnik náhodných jevů**  
Průnik jevů  $A$  a  $B$  ( $A \cap B$ ) nastává, když nastaly jevy  $A$  a  $B$  současně.
- **Jev opačný**  
Jev opačný k jevu  $A$ , značíme  $\bar{A}$ , nastane tehdy, když nenastane jev  $A$ . Někdy označujeme  $\bar{A}$  jako doplněk jevu  $A$ .

Jevy  $A$  a  $B$  nazýváme **disjunktní** (neslučitelné), nemohou-li nastat současně.

**Příklad 1.** *Mějme náhodný pokus jako hod klasickou šestistěnnou kostkou. Uvažujme jevy  $\Omega$ : padne číslo 1, 2, 3, 4, 5 nebo 6,  $N$ : nepadne nic,  $A$ : padne číslo 1 nebo 2 a  $B$ : padne sudé číslo. Pojmenujte jevy  $\Omega$  a  $N$ , určete sjednocení a průnik  $A$  a  $B$  a určete  $\bar{A}$ .*

$$\begin{aligned}\Omega &= \{1, 2, 3, 4, 5, 6\} - \text{jev jistý}, N = \emptyset - \text{jev nemožný}, \\ A &= \{1, 2\}, B = \{2, 4, 6\}, A \cup B = \{1, 2, 4, 6\}, A \cap B = \{2\}, \bar{A} = \{3, 4, 5, 6\}.\end{aligned}$$

**Definice 1.** *Pravděpodobnostní prostor je trojice  $(\Omega, \mathcal{A}, P)$ , kde*

- $\Omega$  je neprázdná množina všech elementárních jevů  $\omega \in \Omega$ ,
- $\mathcal{A}$  ( $\sigma$ -algebra) je takový systém podmnožin  $\Omega$ , že
  - $\Omega \in \mathcal{A}$ ,
  - $\forall A \in \mathcal{A} : \bar{A} \in \mathcal{A}$  - uzavřenost vůči doplňku,
  - $(\forall n \in \mathbb{N} : A_n \in \mathcal{A}) \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$  -  $\sigma$ -aditivita,
- $P$  (pravděpodobnost) je funkce  $\mathcal{A} \rightarrow \langle 0, 1 \rangle$  taková, že  $\forall A \in \mathcal{A}$  platí
  - $P(A) \geq 0$ ,
  - $P(\Omega) = 1$  - pravděpodobnost jevu jistého je 1,
  - $A_1, A_2, \dots \in \mathcal{A}, \forall i \neq j : A_i \cap A_j = \emptyset \Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ .

*Prvky množiny  $\mathcal{A}$  nazýváme náhodnými jevy.*

Pravděpodobnost je tedy číslo popisující míru očekávatelnosti uskutečnění nějakého náhodného jevu. Pro úplnost si uvedme ještě několik důležitých a potřebných vlastností pravděpodobnosti:

- $P(\emptyset) = 0$  - pravděpodobnost jevu nemožného je 0,
- $P(\bar{A}) = 1 - P(A)$ ,
- $A \subset B \Rightarrow P(A) \leq P(B)$ ,
- $P(B - A) = P(B) - P(A \cap B)$ ,  
speciálně  $A \subset B \Rightarrow P(B - A) = P(B) - P(A)$ ,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ,  
speciálně  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ ,
- $P(A) = P(A \cap B) + P(A \cap \bar{B})$ ,
- následně z de Morganových zákonů plyne,
  - $P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B})$ ,
  - $P(A \cap B) = 1 - P(\overline{A \cap B}) = 1 - P(\bar{A} \cup \bar{B})$ .

## 2.2 Nezávislost jevů, podmíněná pravděpodobnost

V následujících řádcích předpokládejme vždy, že zmíněné náhodné jevy jsou ze stejného pravděpodobnostního prostoru  $(\Omega, \mathcal{A}, P)$ . Pokud o nějakém jevu víme, že nastal v případě, kdy nastal nějaký jev jiný, umíme vyjádřit podmíněnou pravděpodobnost těchto jevů.

**Definice 2.** *Mějme náhodné jevy  $A$  a  $B$ , přičemž  $P(B) > 0$ . Podmíněnou pravděpodobnost jevu  $A$  za podmínky, že nastal jev  $B$ , definujeme jako*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1)$$

Osamostatněním průniku jevů  $A$  a  $B$  získáme vztahy

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A), \quad \text{je-li } P(A) > 0. \quad (2)$$

Je nutno však rozlišovat pojmy jevy nezávislé a disjunktní, mezi těmito pojmy neplatí ekvivalence. Pro dva náhodné jevy může platit, že jsou-li nezávislé, nemusí být disjunktní.

**Definice 3.** *Mějme náhodné jevy  $A$  a  $B$ . Nazveme je jevy nezávislými, platí-li vztah*

$$P(A \cap B) = P(A)P(B). \quad (3)$$

**Definice 4.** *Mějme náhodné jevy  $A$ ,  $B$  a  $C$ , kde  $P(C) > 0$ . Jevy  $A$  a  $B$  nazveme podmíněně nezávislé za podmínky, že nastal jev  $C$ , platí-li vztah*

$$P(A \cap B|C) = P(A|C)P(B|C). \quad (4)$$

**Definice 5.** *Jevy  $A_1, A_2, \dots, A_n$  nazveme (vzájemně) nezávislé, když pro každou neprázdnou podmnožinu  $Q \subseteq \{A_1, A_2, \dots, A_n\}$  platí*

$$P\left(\bigcap_{A \in Q} A\right) = \prod_{A \in Q} P(A).$$

**Definice 6.** *Jevy  $A_1, A_2, \dots, A_n$  nazveme (vzájemně) podmíněně nezávislé za podmínky, že nastal jev  $C$ , kde  $P(C) > 0$ , když pro každou neprázdnou podmnožinu  $Q \subseteq \{A_1, A_2, \dots, A_n\}$  platí*

$$P\left(\bigcap_{A \in Q} A|C\right) = \prod_{A \in Q} P(A|C).$$

Je zřejmé, že uvažujeme-li dva nezávislé jevy  $A$  a  $B$ , tak pravděpodobnost jevu  $A$  za podmínky, že nastal jev  $B$ , je rovna

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$



**Věta 1.** Pro průnik  $n$  náhodných jevů  $A_1, A_2, \dots, A_n$  takových, že  $P(\bigcap_{i=1}^{n-1} A_i) > 0$ , platí tzv. řetězové pravidlo

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{n-1} A_i). \quad (5)$$

*Důkaz.* Provedeme důkaz matematickou indukcí. Pro  $n = 1$  věta zjevně platí, protože pro každý jev platí, že  $P(A) = P(A)$ , tedy  $P(A_1) = P(A_1)$ .

Indukční předpoklad: Předpokládejme, že tvrzení platí pro  $n = k$ . že tvrzení platí pro  $n = k$ . Pak ukážeme, že tvrzení platí i pro  $n = k + 1$ . Do výpočtu nám přibyl nový jev  $A_{k+1}$ , průnik všech jevů je tedy průnik prvních  $k$  jevů s jevem s indexem  $k + 1$ .

$$P\left(\bigcap_{i=1}^{k+1} A_i\right) = P(A_{k+1} \cap \bigcap_{i=1}^k A_i).$$

Pak podle vztahu 2 můžeme tvrdit, že

$$P\left(\bigcap_{i=1}^{k+1} A_i\right) = P(A_{k+1}|\bigcap_{i=1}^k A_i)P\left(\bigcap_{i=1}^k A_i\right).$$

A použitím indukčního předpokladu získáváme

$$P\left(\bigcap_{i=1}^{k+1} A_i\right) = P(A_{k+1}|\bigcap_{i=1}^k A_i)P(A_k|\bigcap_{i=1}^{k-1} A_i)P(A_{k-1}|\bigcap_{i=1}^{k-2} A_i) \cdots P(A_2|A_1)P(A_1).$$

Což jsme chtěli dokázat. Věta platí. □

### 2.3 Věta o úplné pravděpodobnosti, Bayesova věta

V předchozí části jsme si vysvětlili potřebný základ pro následující práci s podmíněnými pravděpodobnostmi a závislostmi náhodných jevů. Ten v této kapitole využijeme k formulaci několika důležitých vět teorie pravděpodobnosti. V definicích a větách dále předpokládáme, že je dán pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ .

**Definice 7.** Úplným systémem vzájemně disjunktních jevů nazveme takové jevy  $B_1, B_2, \dots, B_n$ , pro které platí

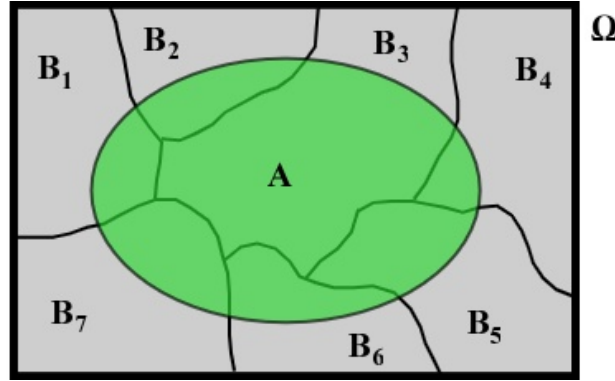
- $\forall i, j \in \{1, 2, \dots, n\} : i \neq j \Rightarrow B_i \cap B_j = \emptyset,$
- $\bigcup_{i=1}^n B_i = \Omega.$

**Věta 2 (Věta o úplné pravděpodobnosti).** Mějme systém vzájemně disjunktních jevů  $B_1, B_2, \dots, B_n$  a jev  $A$  s předpokladem  $P(B_i) > 0, i = 1, 2, \dots, n$ . Pak platí

$$P(A) = \sum_{i=1}^n P(B_i)P(A|B_i). \quad (6)$$

*Důkaz.* Stačí si uvědomit, že pravděpodobnost jevu  $A$  je součet pravděpodobností průniků jevu  $A$  s dílčími jevy  $B_i$  a použít vztah (2):  $P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(B_i)P(A|B_i).$  □

Díky znalosti věty o úplné pravděpodobnosti jsme schopni jednoduše odvodit vztah pro výpočet podmíněné pravděpodobnosti jakéhokoliv náhodného jevu z určitého úplného systému vzájemně disjunktních jevů za podmínky, že nastal jev  $A$ .



Obrázek 1: Ukázka úplného systému vzájemně disjunktních jevů a jeho průnik s jevem A

**Věta 3 (Bayesova věta).** *Mějme systém vzájemně disjunktních jevů  $B_1, B_2, \dots, B_n$  a jev A s předpoklady  $P(A) > 0, P(B_i) > 0, i = 1, 2, \dots, n$ . Pak platí*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)}. \quad (7)$$

*Důkaz.* Důkaz je triviální s použitím předchozí věty, tj. věty o úplné pravděpodobnosti.  $P(B_j|A)$  lze ze vztahu (1) zapsat jako

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}.$$

Čitatel upravíme dle vztahu (2) a jmenovatel podle již zmíněné věty o úplné pravděpodobnosti (7) a získáváme výsledný vztah

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)}.$$

□

**Příklad 2.** *V populaci se vyskytla zákeřná nemoc SARS-CoV-2. Byl vyvinut test, který s pravděpodobností 99 % správně označí nakaženého člověka za pozitivního a s pravděpodobností 5 % je falešně pozitivní. Podle odhadu odborníků je nakaženo 20 % populace. Určete pravděpodobnost, že náhodně testovaný člověk je opravdu nakažený, má-li pozitivní výsledek testu. (Data jsou pouze ilustrační)*

*Označíme-li náhodné jevy*

*B - testovaný je nakažen,*

*A - testovaný má pozitivní test,*

*pak  $P(B) = 0.2 \Leftrightarrow P(\bar{B}) = 0.8, P(A|B) = 0.99$  a  $P(A|\bar{B}) = 0.05$ . Protože B a  $\bar{B}$  tvoří úplný systém vzájemně disjunktních jevů, můžeme dosadit do Bayesovy věty (7)*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} = \frac{0.99 \cdot 0.2}{0.99 \cdot 0.2 + 0.05 \cdot 0.8} \approx 83.19 \ %.$$

*Pravděpodobnost, že testovaný je nakažen, má-li pozitivní výsledek testu je zhruba 83.19 %.*

### 3 Náhodná veličina, náhodný vektor

#### 3.1 Náhodná veličina a její diskretní a absolutně spojitě rozdělení

Uvažujme funkci, která každému elementárnímu jevu z nějakého jevového prostoru přiřazuje reálné číslo. Takovou funkci nazýváme náhodnou veličinou. Mějme pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ .

**Definice 8.** Náhodnou veličinou nazveme funkci  $X : \Omega \rightarrow \mathbb{R}$  takovou, že  $(\forall x \in \mathbb{R})$

$$\{\omega \in \Omega, X(\omega) < x\} \in \mathcal{A}.$$

Náhodné veličiny značíme zpravidla velkými písmeny z konce latinské abecedy, přičemž její jednotlivé realizace značíme písmeny malými. Je-li  $X$  náhodná veličina, zápisem  $X < x$  rozumíme náhodný jev  $\{\omega \in \Omega : X(\omega) < x\}$ . Obdobné značení používáme pro ostatní typy nerovností.

**Definice 9.** Distribuční funkci  $F_X(x)$  náhodné veličiny  $X$  rozumíme funkci  $\mathbb{R} \rightarrow \langle 0, 1 \rangle$  takovou, že  $(\forall x \in \mathbb{R})$

$$F_X(x) = P(X \leq x). \quad (8)$$

Za zmínku stojí některé vlastnosti distribuční funkce  $F_X(x)$  definované vztahem (8):

- $F_X(x)$  je zprava spojitá neklesající funkce,
- $F_X(x)$  nabývá hodnot z intervalu  $\langle 0, 1 \rangle$ , přičemž  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  a  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ ,
- $P(a < X \leq b) = F_X(b) - F_X(a)$  pro  $a, b \in \mathbb{R}; a < b$ ,
- $P(X > a) = 1 - F_X(a)$  pro  $a \in \mathbb{R}$ .

Rozlišujeme dva nejčastější typy náhodných veličin, a to diskretní a absolutně spojitě.

**Definice 10.** Náhodnou veličinu  $X$  nazýváme diskretní (má diskretní rozdělení) jestliže existuje nejvýše spočetná množina  $\mathcal{X} \subset \mathbb{R}$  taková, že

$$\sum_{x \in \mathcal{X}} P(X = x) = 1.$$

Funkci  $p_X(x) = P(X = x)$  pak nazýváme pravděpodobnostní funkcí náhodné veličiny  $X$ . Jako příklady takové náhodné veličiny můžeme uvést

- počet ok při hodu šestistěnnou hrací kostkou,
- počet bodů ve zkoušce z Obyčejných diferenciálních rovnic,
- součet žlutých karet všech hráčů týmu ve sportovním utkání.

Distribuční funkce v bodě  $x$  diskretní náhodné veličiny  $X$  lze vyjádřit jako součet hodnot pravděpodobnostních funkcí v bodech menších než nebo rovných  $x$ , tedy

$$F_X(x) = \sum_{\tilde{x} \in \mathcal{X}: \tilde{x} \leq x} P(X = \tilde{x}).$$

**Definice 11.** Mějme náhodnou veličinu  $X$ . Pokud existuje funkce  $f : \mathbb{R} \rightarrow \langle 0, \infty \rangle$ , taková, že  $(\forall x \in \mathbb{R})$

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad (9)$$

pak náhodnou veličinu  $X$  nazýváme absolutně spojitou ( $X$  má absolutně spojitě rozdělení) a funkci  $f_X(x)$  nazýváme hustotou pravděpodobnosti náhodné veličiny  $X$ .

Pro každou hustotu pravděpodobnosti náhodné veličiny  $X$  platí, že

$$\int_{-\infty}^{\infty} f_X(x) dx = \lim_{t \rightarrow \infty} \int_{-\infty}^t f_X(x) dx = \lim_{t \rightarrow \infty} F_X(t) = 1.$$

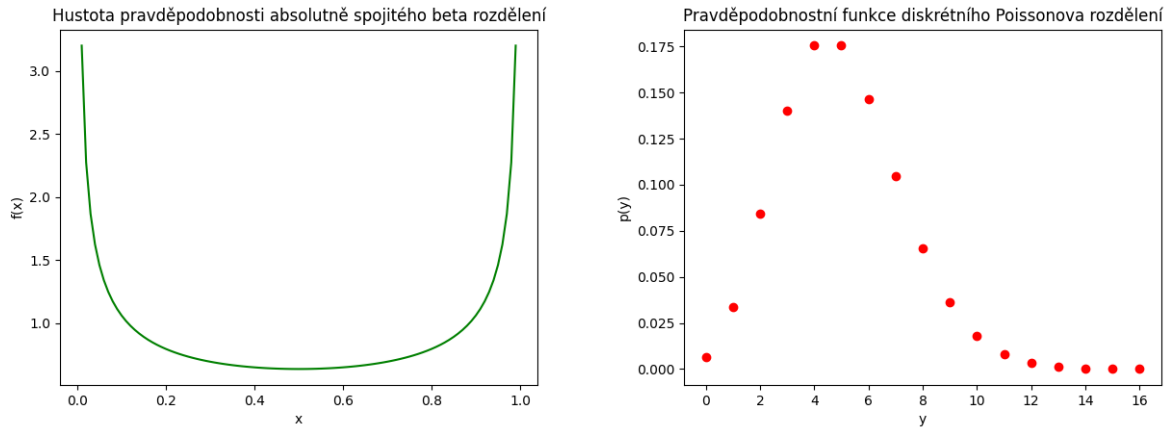
Pravděpodobnostní funkce absolutně spojitě náhodné veličiny je v každém bodě rovna nule. Z definic distribuční funkce a hustoty pravděpodobnosti plynou následující vztahy pro náhodné veličiny s absolutně spojitým rozdělením. Pro  $a, b, x \in \mathbb{R}$  a  $a < b$  platí:

- $P(X = a) = 0$ ,
- $P(X < a) = P(X \leq a) = \int_{-\infty}^a f_X(x) dx$ ,
- $P(X > a) = P(X \geq a) = 1 - P(X < a) = 1 - F_X(a) = 1 - \int_{-\infty}^a f_X(x) dx = \int_a^{\infty} f_X(x) dx$ ,
- $P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^b f_X(x) dx - \int_{-\infty}^a f_X(x) dx = \int_a^b f_X(x) dx$ .

Jako příklady absolutně spojitých náhodných veličin můžeme uvést

- výsledek měření proudu v obvodu,
- hmotnost člověka,
- doba do poruchy zařízení.

Na obrázku (2) vidíme vlevo hustotu pravděpodobnosti  $f_X(x)$  absolutně spojitě náhodné veličiny  $X$  s beta rozdělením s parametry  $\alpha = 0.5$ ,  $\beta = 0.5$  a vpravo pravděpodobnostní funkci  $p_Y(y)$  diskrétní náhodné veličiny  $Y$  s Poissonovým rozdělením s parametrem  $\lambda = 5$ .



Obrázek 2: Ilustrace hustoty pravděpodobnosti a pravděpodobnostní funkce

### 3.2 Náhodný vektor

Mnohdy je třeba pracovat i s více než jednou náhodnou veličinou. Jako zobecnění náhodné veličiny v tomto smyslu definujeme náhodný vektor. Uvažujme, že se následující řádky týkají daného pravděpodobnostního prostoru  $(\Omega, \mathcal{A}, P)$  a  $n \in \mathbb{N}$ .

**Definice 12.**  *$n$ -rozměrným náhodným vektorem nazveme uspořádanou  $n$ -tici náhodných veličin  $X_1, X_2, \dots, X_n$ . Značíme je  $X = (X_1, X_2, \dots, X_n)$ .*

Náhodná veličina je tedy speciálním případem náhodného vektoru dimenze 1, je tedy zřejmé, že následující definice budou v jistém smyslu rozšířením definic z předchozí sekce s analogickým významem.

**Definice 13.** *Sdruženou (simultánní) distribuční funkcí  $F_X(x_1, \dots, x_n)$  náhodného vektoru  $X = (X_1, \dots, X_n)$  rozumíme funkci  $\mathbb{R}^n \rightarrow \langle 0, 1 \rangle$  takovou, že*

$$F_X(x) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Pro sdruženou distribuční funkci  $F_X(x_1, \dots, x_n)$  náhodného vektoru  $X = (X_1, \dots, X_n)$  platí následující vlastnosti

- $F_X(x_1, \dots, x_n)$  je v každé proměnné zprava spojitá neklesající funkce,
- $F_X(x_1, \dots, x_n)$  nabývá hodnot z intervalu  $\langle 0, 1 \rangle$ , přičemž  
 $\forall i \in \{1, \dots, n\} : \lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_n) = 0$  a  $\lim_{(x_1, \dots, x_n) \rightarrow (+\infty, \dots, +\infty)} F_X(x_1, \dots, x_n) = 1$ .

Obdobně jako u náhodných veličin dělíme náhodný vektor na dva základní typy, a to náhodné vektory diskrétní a absolutně spojitý.

**Definice 14.** *Náhodný vektor  $X = (X_1, \dots, X_n)$  nazýváme diskrétní (má diskrétní rozdělení) jestliže existují nejvýše spočetné množiny  $(\mathcal{X}_1, \dots, \mathcal{X}_n) \subset \mathbb{R}^n$  takové, že*

$$\sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_n \in \mathcal{X}_n} P(X_1 = x_1, \dots, X_n = x_n) = 1.$$

Funkci  $p_X(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$  pak nazýváme sdruženou pravděpodobnostní funkcí náhodného vektoru  $X$ .

**Definice 15.** *Mějme náhodný vektor  $X = (X_1, \dots, X_n)$ . Pokud existuje funkce  $f : \mathbb{R}^n \rightarrow \langle 0, \infty \rangle$  taková, že*

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n) dt_n \dots dt_1, \quad (10)$$

*pak náhodný vektor  $X$  nazýváme absolutně spojitý ( $X$  má absolutně spojitě rozdělení) a funkci  $f_X(x_1, \dots, x_n)$  nazýváme sdruženou hustotou pravděpodobnosti náhodného vektoru  $X$ .*

Pro sdruženou hustotu pravděpodobnosti náhodného vektoru  $X$  dále platí

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(x_1, \dots, x_n) dx_n \dots dx_1 = 1.$$

Pro vyjádření rozdělení některých ze složek náhodného vektoru používáme marginální rozdělení, tj. marginální distribuční a pravděpodobnostní funkci nebo marginální hustotu pravděpodobnosti náhodného vektoru.

Uvažujme náhodný vektor  $X = (X_1, \dots, X_n)$  a  $i_1, \dots, i_k \in \{1, \dots, n\}$ ,  $k < n$  jako navzájem různé indexy. Pak libovolný náhodný vektor  $Y = (X_{i_1}, \dots, X_{i_k})$  označme jako "podvektor" náhodného vektoru  $X$ . Dále existuje  $n - k$  indexů  $j_1, \dots, j_{n-k} \in \{1, \dots, n\}$  takových, že jsou různé od indexů  $i_1, \dots, i_k$  a spolu s těmito indexy tvoří dohromady celou množinu  $\{1, \dots, n\}$ . Náhodný vektor  $Z = (X_{j_1}, \dots, X_{j_{n-k}})$  je pak také podvektorem a jeho složky spolu se složkami náhodného vektoru  $Y$  tvoří dohromady původní náhodný vektor  $X$ .

**Definice 16.** *Nechť  $X = (X_1, \dots, X_n)$  je náhodný vektor a  $Y = (X_{i_1}, \dots, X_{i_k})$ ,  $i_1, \dots, i_k \in \{1, \dots, n\}$ ,  $k < n$  nějaký jeho podvektor. Pak distribuční funkci  $F_Y(x_{i_1}, \dots, x_{i_k})$  říkáme marginální distribuční funkce podvektoru  $Y$ .*

Marginální distribuční funkci podvektoru  $Y$  pak můžeme vyjádřit

$$F_Y(x_{i_1}, \dots, x_{i_k}) = \lim_{(x_{j_1}, \dots, x_{j_{n-k}}) \rightarrow (+\infty, \dots, +\infty)} F_X(x_1, \dots, x_n). \quad (11)$$

**Definice 17.** *Nechť  $X = (X_1, \dots, X_n)$  je diskrétní náhodný vektor s hodnotami v  $\mathcal{X}_1 \times \dots \times \mathcal{X}_n \subset \mathbb{R}^n$  a  $Y = (X_{i_1}, \dots, X_{i_k})$ ,  $i_1, \dots, i_k \in \{1, \dots, n\}$ ,  $k < n$  nějaký jeho podvektor. Pak pravděpodobnostní funkci  $p_Y(x_{i_1}, \dots, x_{i_k})$  říkáme marginální pravděpodobnostní funkce podvektoru  $Y$ .*

Pro marginální pravděpodobnostní funkci podvektoru  $Y$  pak platí

$$p_Y(x_{i_1}, \dots, x_{i_k}) = \sum_{x_{j_1} \in \mathcal{X}_{j_1}} \dots \sum_{x_{j_{n-k}} \in \mathcal{X}_{j_{n-k}}} p(x_1, \dots, x_n). \quad (12)$$

**Definice 18.** *Nechť  $X = (X_1, \dots, X_n)$  je náhodný vektor a  $Y = (X_{i_1}, \dots, X_{i_k})$ ,  $i_1, \dots, i_k \in \{1, \dots, n\}$ ,  $k < n$  nějaký jeho podvektor. Pak hustotě pravděpodobnosti  $f_Y(x_{i_1}, \dots, x_{i_k})$  říkáme marginální hustota pravděpodobnosti podvektoru  $Y$ .*

Marginální hustotu podvektoru  $Y$  pak můžeme obdobně vyjádřit

$$f_Y(x_{i_1}, \dots, x_{i_k}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_{j_{n-k}} \dots dx_{j_1}. \quad (13)$$

Velice důležitým pojmem je pak nezávislost složek náhodného vektoru.

**Definice 19.** *Mějme náhodný vektor  $X = (X_1, \dots, X_n)$ . Jeho složky  $X_1, \dots, X_n$  nazveme nezávislými, platí-li vztah*

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i). \quad (14)$$

Tuto definici můžeme separovat podle typu náhodného vektoru. Složky náhodného vektoru  $X = (X_1, \dots, X_n)$  nazveme nezávislými, platí-li

- $p_X(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$  pro diskrétní náhodný vektor  $X$ ,
- $f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$  pro absolutně spojitý náhodný vektor  $X$ .

Dalšími esenciálními pojmy jsou podmíněná pravděpodobnostní funkce a podmíněná hustota pravděpodobnosti. Pro jednodušší zavedení se omezme pouze na dvojrozměrný (respektive třírozměrný) náhodný vektor na daném pravděpodobnostním prostoru.

**Definice 20.** Mějme náhodné veličiny  $X$  a  $Y$ . Podmíněnou pravděpodobnostní funkci  $X$  za podmínky, že nastal jev  $Y = y$ , definujeme jako

$$p_{X|Y}(x, y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad p_Y(y) \neq 0. \quad (15)$$

**Definice 21.** Mějme náhodné veličiny  $X$  a  $Y$ . Podmíněnou hustotu pravděpodobnosti  $X$  za podmínky, že nastal jev  $Y = y$ , definujeme jako

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad f_Y(y) \neq 0. \quad (16)$$

Jsou-li dvě náhodné veličiny nezávislé, je snadné si povšimnout, že

- $p_{X|Y}(x, y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = p_X(x)$  pro diskrétní  $X$  a  $Y$ ,
- $f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = f_X(x)$  pro absolutně spojitě  $X$  a  $Y$ .

Uvažujme nyní náhodné veličiny  $X$ ,  $Y$  a  $Z$ . Využitím věty (5) můžeme tvrdit, že platí

- $p_{X,Y,Z}(x, y, z) = p_{X|Y,Z}(x|y, z)p_{Y|Z}(y|z)p_Z(z)$  pro diskrétní  $X$ ,  $Y$  a  $Z$ ,
- $f_{X,Y,Z}(x, y, z) = f_{X|Y,Z}(x|y, z)f_{Y|Z}(y|z)f_Z(z)$  pro absolutně spojitě  $X$ ,  $Y$  a  $Z$ .

**Definice 22.** Mějme náhodné veličiny  $X$ ,  $Y$  a  $Z$ .  $X$  a  $Y$  nazveme nezávislými za podmínky, že nastal jev  $Z = z$ , platí-li vztah

$$F_{X,Y|Z}(x, y|z) = F_{X|Z}(x|z)F_{Y|Z}(y|z). \quad (17)$$

Opět můžeme podmíněnou nezávislost náhodných veličin separovat podle jejich typu. Náhodné veličiny  $X$  a  $Y$  nazveme nezávislými za podmínky, že nastal jev  $Z = z$ , platí-li

- $p_{X,Y|Z}(x, y|z) = p_{X|Z}(x|z)p_{Y|Z}(y|z)$  pro diskrétní náhodné veličiny,
- $f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z)$  pro absolutně spojitě náhodné veličiny.

Z čehož vychází následující vlastnost pravděpodobnostní funkce a hustoty pravděpodobnosti

- $p_{X|Y,Z}(x|y, z) = p_{X|Z}(x|z)$  pro diskrétní náhodné veličiny,
- $f_{X|Y,Z}(x|y, z) = f_{X|Z}(x|z)$  pro absolutně spojitě náhodné veličiny.

Bayesovu větu (7) můžeme aplikovat i na náhodné vektory, respektive jejich složky. Dostáváme pak pro diskrétní náhodné veličiny  $X$  (s hodnotami v  $\mathcal{X} \subset \mathbb{R}$ ) a  $Y$  vztah

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)} = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{x \in \mathcal{X}} p_{Y|X}(y|x)p_X(x)}, \quad (18)$$

a pro náhodné veličiny absolutně spojitě pak

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x) dx}. \quad (19)$$

### 3.3 Číselné charakteristiky náhodné veličiny a náhodného vektoru

Někdy se hodí shrnout informace o nějaké náhodné veličině, respektive náhodném vektoru, do jednoho údaje, se kterým se pak snáze pracuje. Takovým údajům říkáme číselné charakteristiky. Definice se však často liší pro diskrétní a absolutně spojitou náhodnou veličinu, principiálně mají ale stejný význam.

**Definice 23.** Střední hodnota  $E(X)$  náhodné veličiny  $X$  je reálné číslo takové, že

$$E(X) = \begin{cases} \sum_{x \in \mathcal{X}} xp(x) & \text{v případě diskrétní náhodné veličiny,} \\ \int_{-\infty}^{\infty} xf(x) dx & \text{v případě absolutně spojitě náhodné veličiny,} \end{cases} \quad (20)$$

pokud příslušné řady, resp. integrály konvergují absolutně<sup>1</sup>.

Střední hodnota je tedy nejzákladnější číselnou charakteristikou polohy náhodné veličiny. Uvedme si některé její vlastnosti:

- $E(aX + b) = aE(X) + b$  pro všechna  $a, b \in \mathbb{R}$ ,
- $E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$ ,
- $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$ , jestliže jsou náhodné veličiny  $X_i$  vzájemně nezávislé.

**Definice 24.** Střední hodnotou náhodného vektoru  $X = (X_1, \dots, X_n)$  rozumíme vektor středních hodnot jeho složek:  $E(X) = [E(X_1), \dots, E(X_n)]$ .

**Definice 25.** Rozptyl  $D(X)$  náhodné veličiny  $X$  je nezáporné reálné číslo takové, že

$$D(X) = E([X - E(X)]^2). \quad (21)$$

Reálné číslo  $\sqrt{D(X)} = \sigma$  pak nazýváme směrodatnou odchylkou náhodné veličiny  $X$ .

Mezi nejdůležitější vlastnosti rozptylu patří:

- $D(aX + b) = a^2 D(X)$  pro všechna  $a, b \in \mathbb{R}$ ,
- $D(X) = E(X^2) - [E(X)]^2$ ,
- $D(\sum_{i=1}^n X_i) = \sum_{i=1}^n D(X_i)$ , jestliže jsou náhodné veličiny  $X_i$  vzájemně nezávislé.

**Definice 26.** Kovarianci náhodných veličin  $X$  a  $Y$  definujeme jako

$$Cov(X, Y) = E([X - E(X)][Y - E(Y)]). \quad (22)$$

Z definic kovariance a rozptylu vyplývá, že  $D(X) = Cov(X, X)$ . Dále je vhodné zmínit, že

- pro nezávislé náhodné veličiny  $X$  a  $Y$  platí  $Cov(X, Y) = 0$ ,
- $Cov(X, Y) = E(XY) - E(X)E(Y)$ .

---

<sup>1</sup>Řadu  $\sum_{n=1}^{\infty} a_n$ , existuje-li konečná  $\lim s_n$ , kde  $s_n = \sum_{k=1}^n a_k$ , nazýváme konvergentní. Konverguje-li i řada  $\sum_{n=1}^{\infty} |a_n|$ , řadu  $\sum_{n=1}^{\infty} a_n$  nazýváme absolutně konvergentní. Nevlastní integrál (pokud má tento integrál smysl)  $\int_a^{\infty} f(x) dx$ , existuje-li konečná  $\lim_{c \rightarrow \infty} \int_a^c f(x) dx$ , nazýváme konvergentní. Konverguje-li i integrál  $\int_a^{\infty} |f(x)| dx$ , integrál  $\int_a^{\infty} f(x) dx$  nazýváme absolutně konvergentní.



### 3.4 Normální (Gaussovo) rozdělení náhodné veličiny

Normální rozdělení je velice důležitým pravděpodobnostním rozdělením. Popisuje mnoho reálných veličin (výšku člověka, chyby fyzikálních měření) a centrální limitní věta<sup>2</sup> dokonce zhruba tvrdí, že za pomoci normálního rozdělení dokážeme za vhodných podmínek aproximovat například součet většího množství libovolných nezávislých stejně rozdělených náhodných veličin. Své uplatnění získává také v pravděpodobnostních modelech.

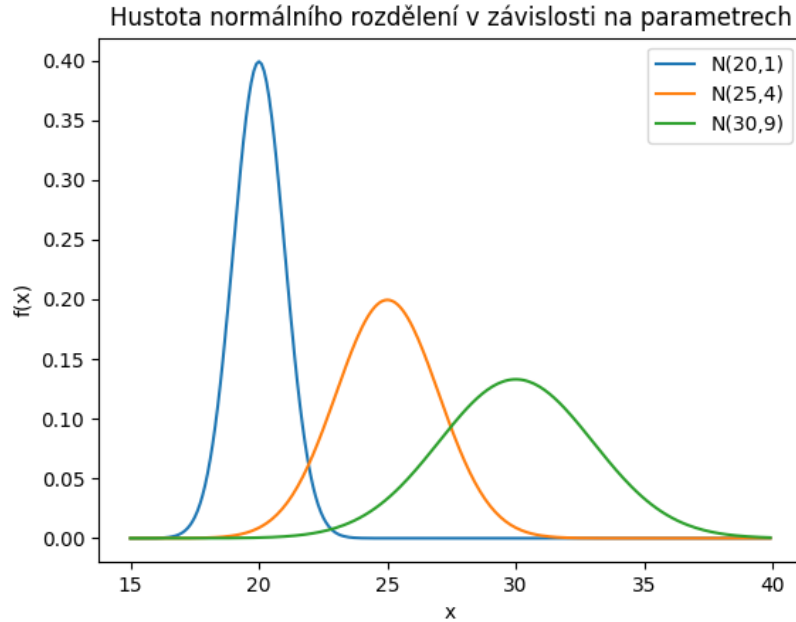
**Definice 27.** Má-li náhodná veličina  $X$  hustotu pravděpodobnosti

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}, \quad (23)$$

pak řekneme, že má normální rozdělení  $X \sim N(\mu, \sigma^2)$ , kde  $\mu$  a  $\sigma^2$ ,  $\sigma^2 > 0$  jsou reálné parametry tohoto rozdělení.

Pro  $X \sim N(\mu, \sigma^2)$  platí, že  $E(X) = \mu$  a  $D(X) = \sigma^2$ . Normální rozdělení je tedy jasně určeno svou střední hodnotou  $E(X) = \mu$  a rozptylem  $D(X) = \sigma^2$ , tedy jejími parametry. Někdy (zvláště v anglosaské literatuře) se jako druhý parametr normálního rozdělení uvádí místo rozptylu směrodatná odchylka  $\sigma$ .

Hustota pravděpodobnosti má typický tvar zvonovité křivky (anglicky „bell curve“). Na obrázku (4) si pro představu ukažme tvar této křivky pro rozdílné parametry. Modře je vyznačena hustota pro parametry  $\mu_1 = 20$ ,  $\sigma_1^2 = 1$ , oranžově pro  $\mu_2 = 25$ ,  $\sigma_2^2 = 4$  a zeleně pro  $\mu_3 = 30$ ,  $\sigma_3^2 = 9$ .



Obrázek 3: Hustota normálního rozdělení v závislosti na parametrech

<sup>2</sup>Centrální limitní věta (Central limit theorem - CLT) říká, že pro větší počet (v praxi  $> 30$ ) libovolných nezávislých stejně rozdělených náhodných veličin  $X_1, \dots, X_n$  se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$  má výběrový průměr  $\bar{X}_n$  přibližně normální rozdělení:  $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ .

## 4 Markovský řetězec, markovské náhodné pole

### 4.1 Markovský řetězec

Představme si, že se zabýváme souborem nekonečně mnoha náhodných veličin. V tomto smyslu se jako zobecnění náhodného vektoru, typicky pro nespočetně mnoho náhodných veličin, zavádí pojem náhodný proces. Pro následující řádky opět uvažujme pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ . V této kapitole bylo čerpáno především z [3] a [4].

**Definice 28.** *Náhodným procesem  $X$  nazveme soubor náhodných veličin  $X = (X_t)_{t \in T}$ , kde  $X_t$  jsou náhodné veličiny a  $T \subset \mathbb{R}$ .*

Indexy  $t$  se často interpretují jako čas. Náhodné procesy tedy můžeme chápat jako modely pro náhodné funkce, resp. posloupnosti, proto je dle charakteru množiny  $T$  dělíme na náhodné procesy

- se spojitým časem, je-li  $T$  intervalem reálných čísel,
- s diskrétním časem, je-li  $T$  nejvýše spočetná množina.

Dále můžeme náhodné procesy dělit dle typu dílčích náhodných veličin  $X_t$  na náhodné procesy

- se spojitými stavy,
- s diskrétními stavy.

Markovským řetězcem se většinou rozumí markovský proces s diskrétním časem. Množinu  $\{1, 2, \dots, k\}$  všech stavů řetězce označme  $S$ . Pro jednoduchost se budeme zabývat pouze markovskými řetězci s konečným počtem stavů.

**Definice 29.** *Celočíselnou posloupnost  $(X_n)_{n \in \mathbb{N}_0}$  nazveme markovským řetězcem, jestliže platí  $(\forall n \in \mathbb{N})$*

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n), \quad (24)$$

kde  $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) > 0$  a  $i_0, \dots, i_n, j \in S$ .

Pravděpodobnost nastání následujícího stavu je tedy podmíněně nezávislá na minulých stavech, když známe stav přítomný. Této vlastnosti říkáme markovská vlastnost a jde zobecnit do tvaru

$$P(X_{n+m} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_{n+m} = j | X_n = i_n). \quad (25)$$

**Definice 30.** *Podmíněné pravděpodobnosti*

$$P(X_{n+1} = j | X_n = i) \quad (26)$$

říkáme pravděpodobnost přechodu ze stavu  $i$  v čase  $n$  do stavu  $j$  v čase  $n + 1$ .

Pokud pro markovský řetězec platí, že pravděpodobnosti přechodu nejsou závislé na velikosti času, tedy jestliže platí

$$P(X_{n+m+k} = j | X_{m+k} = i) = P(X_{n+m} = j | X_m = i),$$

tak řetězec nazýváme homogenním. V následujícím textu se věnujme jen homogennímu řetězci. Označme pravděpodobnost přechodu ze stavu  $i$  v čase  $n$  do stavu  $j$  v čase  $n + 1$

$$p_{ij} = P(X_{n+1} = j | X_n = i) \quad (27)$$

a pak pravděpodobnosti přechodu

$$p_{ij}(m) = P(X_{n+m} = j | X_n = i) \quad (28)$$

říkáme pravděpodobnost přechodu po  $m$  krocích ( $m$ -tého řádu) ze stavu  $i$  v čase  $n$  do stavu  $j$  v čase  $n + m$  v homogenním markovském řetězci.

Homogenní markovský řetězec se dá popsat maticí, ve které se na pozici  $i, j$  nachází odpovídající pravděpodobnost přechodu  $p_{ij}$ .

**Definice 31.** *Mějme homogenní markovský řetězec. Matici  $\mathbb{P} \in \mathbb{R}^{k \times k}$  takovou, že*

$$\mathbb{P} = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{kk} & \cdots & p_{kk} \end{bmatrix}, \quad (29)$$

*nazýváme maticí pravděpodobností přechodu.*

$\mathbb{P}$  je čtvercová matice s prvky  $p_{ij} \in \langle 0, 1 \rangle$ , která je stochastická, což znamená, že součet prvků v každém řádku je roven 1, tedy

$$\forall i \in S : \sum_{j \in S} p_{ij} = 1. \quad (30)$$

Což vyplývá z faktu, že prvky v každém řádku matice tvoří úplnou soustavu jevů. Matici  $\mathbb{P}(n)$  nazveme maticí pravděpodobností přechodu po  $n$  krocích ( $n$ -tého řádu), jejíž prvky jsou  $p_{ij}(n)$ . Matice pravděpodobností přechodu existují i pro nehomogenní řetězce, mnohé vztahy pro ně však neplatí.

**Věta 4.** *V homogenním markovském řetězci platí Chapmanovy-Kolmogorovy rovnice ve tvaru*

$$p_{ij}(m+n) = \sum_{k \in S} p_{ik}(m)p_{kj}(n), \quad (31)$$

*což se pro matice pravděpodobností přechodu transformuje na tvar*

$$\mathbb{P}(m+n) = \mathbb{P}(m)\mathbb{P}(n). \quad (32)$$

*Důkaz.* Využitím markovské vlastnosti (25) a podmíněné nezávislosti náhodných veličin (17) získáváme

$$\begin{aligned} p_{ij}(m+n) &= P(X_{m+n} = j | X_0 = i) \\ &= \sum_{k \in S} P(X_{m+n} = j, X_m = k | X_0 = i) \\ &= \sum_{k \in S} P(X_{m+n} = j | X_m = k, X_0 = i) P(X_m = k | X_0 = i) \\ &= \sum_{k \in S} P(X_{m+n} = j | X_m = k) P(X_m = k | X_0 = i) \\ &= \sum_{k \in S} p_{kj}(n) p_{ik}(m). \end{aligned}$$

□

V homogenním řetězci tedy zřejmě platí, že  $\mathbb{P}(n) = \mathbb{P}^n$ .

Pro popis markovského řetězce se také používá jeho pravděpodobnostní rozdělení  $p_i(n) = P(X_n = i)$  pro stav  $i$  v čase  $n$ .  $p_i(0) = P(X_0 = i)$  nazýváme počátečním rozdělením markovského řetězce, kde  $i$  označuje počáteční stav.

Rozdělení v daném čase  $n$  se dá popsat vektorem pravděpodobností  $p(n) = (p_i(n))_{i \in S}$ .

**Definice 32.** *Nechť  $(X_n)_{n \in \mathbb{N}_0}$  je markovský řetězec s množinou stavů  $S$  a maticí pravděpodobností přechodu  $\mathbb{P} = (p_{ij})_{i,j \in S}$ , pak rozdělení  $\pi$  nazveme stacionárním, jestliže pro  $i, j \in S$  platí*

$$\pi_j = \sum_{i \in S} p_{ij} \pi_i \quad (33)$$

nebo maticově

$$\pi^T = \pi^T \mathbb{P}. \quad (34)$$

Pro řetězec s konečně mnoha stavy a kladnými pravděpodobnostmi přechodu platí

- $p(n) = \pi \Rightarrow p(n+1) = \pi$ ,
- lze ukázat, že

$$\lim \mathbb{P} = \begin{bmatrix} \pi_1 & \dots & \pi_m \\ \vdots & \ddots & \vdots \\ \pi_1 & \dots & \pi_m \end{bmatrix}, \quad (35)$$

a tedy  $\lim p(0)^T \mathbb{P}^n = \pi^T$  nezávisle na rozdělení počátečního stavu, více v [4].

Stacionární rozdělení popisuje chování markovského řetězce z dlouhodobého hlediska. Pro každý řetězec s konečně mnoha stavy stacionární rozdělení existuje nezávisle na počátečním stavu a pokud se řetězec v čase  $n$  nachází ve stacionárním rozdělení, bude se v něm nacházet i v čase  $n+1$ .

## 4.2 Markovská náhodná pole

**Definice 33.** *Nechť  $S$  je libovolná konečná množina. Náhodný vektor  $X = (X_s)_{s \in S}$ , kde náhodné veličiny  $X_s$  mají hodnoty ve stejných konečných množinách  $\mathcal{X}_s$ , nazveme náhodným polem.*

**Definice 34.** *Systém množin  $\mathcal{N} = (N_s)_{s \in S}$ , kde  $N_s \subset S$  pro který platí*

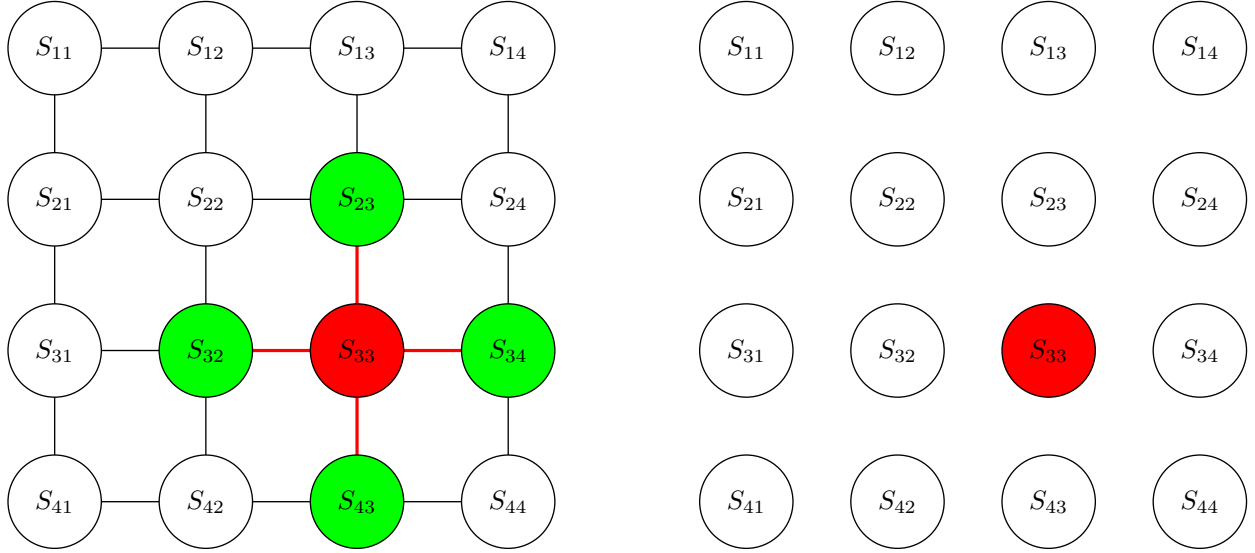
- $\forall s \in S : s \notin N_s$ ,
- $\forall s, t \in S : s \in N_t \Leftrightarrow t \in N_s$ ,

*nazveme systémem okolí  $\mathcal{N}$  a množinu  $C \subset S$  takovou, že  $\forall s, t \in C : s \in N_t$ , nazveme klikou systému  $\mathcal{N}$  a body  $s, t \in S : s \in N_t$  nazýváme sousedy.*

Množinu  $S$  a systém okolí  $\mathcal{N}$  je mnohdy vhodné reprezentovat neorientovaným grafem  $(V, E)$ , kde prvky  $s$  množiny  $S$  jsou vrcholy grafu ( $S = V$ ) a  $E$  je množina hran. Hrany jsou právě mezi těmi vrcholy  $s, t \in S$ , které jsou sousedy.

Často používané systémy okolí jsou například

- $S = \{(i, j) : i \in \{1, \dots, m\}, j \in \{1, \dots, n\}\}$ ,
- $E = \{((i, j), (k, l)) \in S \times S : |i - k| + |j - l| = 1\}$ .  
 $S$  tedy lze chápat jako systém souřadnic  $m \times n$  a pro  $\mathcal{N}$  platí, že každé  $s$  sousedí právě s body přímo nalevo, napravo, nad a pod  $s$ ,
- systém okolí reprezentovaný grafem  $(S, \emptyset)$  (žádné  $s \in S$  nemá souseda).



Obrázek 4: Znázornění dvou systémů okolí  $\mathcal{N}$  neorientovanými grafy

Na obrázku (4) graf vlevo popisuje systém, kdy každý bod má jako sousedy bod přímo nalevo, napravo, pod a nad. Graf vpravo pak popisuje systém, kdy každý bod má právě 0 sousedů. Červeně je vyznačen bod  $s_{33}$ , zeleně jeho sousedé, kteří jsou s bodem  $s_{33}$  spojení červenou hranou. Pro  $M \subset S$  označme  $X_M$  náhodný vektor  $(X_s)_{s \in M}$ .

**Definice 35.** Necht  $X = (X_s)_{s \in S}$  je náhodné pole s rozdělením pravděpodobnosti  $P$  a  $\mathcal{N}$  je systém okolí v  $S$ . Náhodné pole  $X$  nazveme markovským náhodným polem vzhledem k systému okolí  $\mathcal{N}$ , jestliže platí

$$P(x) > 0, \quad (36)$$

$$P(X_s = x_s | X_{S \setminus \{s\}} = x_{S \setminus \{s\}}) = P(X_s = x_s | X_{\mathcal{N}_s} = x_{\mathcal{N}_s}),$$

pro všechna  $x \in \mathcal{X}_s$ ,  $s \in S$ .

Náhodná veličina  $X_s$  v markovském náhodném poli je tedy podmíněně nezávislá na všech ostatních složkách pole, známe-li sousedy bodu  $s$ .

### 4.3 Gibbsovo rozdělení

Gibbsovy distribuce jsou pravděpodobnostní rozdělení, která ve statistické mechanice popisují pravděpodobnostní rozdělení stavů velkého počtu částic.

**Definice 36.** *Nechť  $X$  je náhodné pole s rozdělením pravděpodobnosti  $P$ , které lze vyjádřit ve tvaru*

$$P(X = x) = \frac{\exp(-\beta H(x))}{\sum_{\tilde{x} \in \mathcal{X}} \exp(-\beta H(\tilde{x}))} \quad (37)$$

*pro nějakou funkci  $H : \mathcal{X} \rightarrow \mathbb{R}$ . Rozdělení  $P$  nazveme Gibbsovým rozdělením s energetickou funkcí  $H$ , konstantou  $\beta$  a pole  $X$  nazveme Gibbsovým polem.*

Funkce  $H(x)$  se interpretuje jako energie systému vyjadřující míru interakce mezi složkami pole (často ve smyslu částic, pixelů,...). Ze vztahu (37) lze vyčíst, že pravděpodobnost je klesající funkcí energie  $H(x)$ , takže vyšší pravděpodobnost je přiřazována rozložením s menší energií. Často se energie v Gibbsově poli vyjadřuje jako součet nějakých lokálních příspěvků. Konstanta  $\beta$  určuje význam energie na celkové pravděpodobnosti soustavy.

## 5 Markov Chain Monte Carlo

### 5.1 Monte Carlo metody

V této kapitole bylo čerpáno ze [4] a [5]. Pro jednoduchost se vzhledem k následujícím kapitolám domluvíme, že bude-li význam zřejmý, nebudeme rozlišovat náhodnou veličinu a její hodnotu, a že u hustot pravděpodobností nebudeme indexem rozlišovat příslušnost k dané náhodné veličině, ta je totiž jasná z argumentů daných funkcí. Nakonec předpokládejme, že zmíněné integrály budou, nebude-li řečeno jinak, chápány jako integrály přes celý obor hodnot příslušné náhodné veličiny.

Pro výpočet složitých integrálů (obzvláště mnohdimenzionálních) často nestačí klasické numerické metody a musí se využít simulačních metod, například metody Monte Carlo. Tato třída algoritmů je založena na generování pseudonáhodných čísel, proto je důležitá kvalita generátoru. Tyto metody našly své uplatnění právě i v bayesovské statistice kvůli mnohdy nepřívětivým tvarům aposteriorních hustot. Cílem Monte Carlo metod je tedy přibližný výpočet integrálu

$$I = \int g(x) dx, \quad (38)$$

kde  $g(x)$  je libovolná integrovatelná funkce. Tento integrál můžeme upravit do tvaru

$$I = \int \left[ \frac{g(x)}{f(x)} \right] f(x) dx, \quad (39)$$

kde pro nenulovou funkci  $g(x)$  je  $f(x) > 0$ . Označme pak  $h(x) = \left[ \frac{g(x)}{f(x)} \right]$  a uvažujme náhodnou veličinu  $X$  s hustotou  $f(x)$ , pak integrál  $I$  je roven střední hodnotě  $h(X)$  a získáváme

$$I = E(h(X)) = \int h(x)f(x) dx. \quad (40)$$

Umíme-li generovat nezávislé vzorky  $x^{(1)}, x^{(2)}, \dots$  z rozdělení s hustotou  $f(x)$ , můžeme integrál (58) aproximovat výběrovým průměrem

$$I_N = \frac{1}{N} \sum_{i=1}^N h(x^{(i)}). \quad (41)$$

Pak podle zákona velkých čísel platí  $I_N \xrightarrow{\text{s.j.}} I$  (konvergence skoro jistě<sup>3</sup>).

---

<sup>3</sup>Nechť  $X, X_1, X_2, \dots$  jsou náhodné veličiny na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . Řekneme, že posloupnost  $X_1, X_2, \dots$  konverguje k náhodné veličině  $X$  skoro jistě, jestliže platí  $P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$ .

## 5.2 Markov Chain Monte Carlo metody

Pro klasické Monte Carlo metody je ve více dimenzích obtížné nalézt vhodnou  $f(x)$  (39), proto jsou v těchto případech prakticky nepoužitelné. Pak, zejména v bayesovské statistice, slouží výborně metody Markov Chain Monte Carlo (dále MCMC). Tyto metody negenerují nezávislé vzorky, nýbrž vzorky z Markovova řetězce, jehož stacionární rozdělení  $\pi$  je rovno rozdělení cílovému.

Metropolisův Hastingsův algoritmus je obecný MCMC algoritmus, který iteračně generuje posloupnost vzorků jakožto hodnoty Markovova řetězce  $x^{(1)}, x^{(2)} \dots$ . Vstupem algoritmu jsou

- cílová hustota  $f(x)$  - simulovaná hustota,
- návrhová hustota  $q(x^*|x)$  - hustota generující návrhy,
- počet vzorků  $N$ .

Algoritmus se dá tedy popsat nějak takto:

---

**Algorithm 1:** Metropolisův Hastingsův algoritmus

---

```

zvol  $x^{(0)}$  tak, že  $f(x^{(0)}) > 0$ ;
for  $i$  from 1 to  $N$  do
    vygeneruj  $u \sim U(\langle 0, 1 \rangle)$ ;
    vygeneruj  $x^* \sim q(x^*|x^{(i-1)})$ ;
     $\mathcal{A}(x^{(i-1)}, x^*) = \min \left\{ \frac{f(x^*)q(x^{(i-1)}|x^*)}{f(x^{(i-1)})q(x^*|x^{(i-1)})}, 1 \right\}$ ;
    if  $u < \mathcal{A}(x^{(i-1)}|x^*)$  then
        |  $x^{(i)} = x^*$ ;
    else
        |  $x^{(i)} = x^{(i-1)}$ ;
    end
end

```

---

Zvolme nějaké  $x^{(0)}$  takové, že cílová hustota je v tomto bodě kladná. Což znamená, že i hustota ve všech dalších vzorcích bude rovněž kladná. Následuje cyklus délky odpovídající počtu generovaných vzorků. V první části každé iterace se vygeneruje náhodný vzorek  $u$  z uniformního rozdělení na intervalu  $\langle 0, 1 \rangle$ , vygeneruje se návrh dalšího vzorku  $x^*$  z návrhové hustoty a vypočítá se pravděpodobnost přijetí  $\mathcal{A}$ . Pokud je náhodně vygenerované  $u$  menší než  $\mathcal{A}$ , tak  $x^*$  přijímáme jako nový vzorek a, jinak pro následující iteraci zanecháme aktuální hodnotu, vzorek se nepřijímá. Následuje další iterace.

V případech, že má cílová hustota specifické vlastnosti, se používají speciální verze Metropolisova Hastingsova algoritmu. Jedna z nich se nazývá Gibbsův sampler a používá se, když umíme snadno generovat vzorky z podmíněných hustot ve tvaru

$$f(x) = f(x_j | x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n), \quad (42)$$

které se zvláště v případě například markovovských náhodných polí (58) dají díky podmíněné nezávislosti zjednodušit na mnohem příjemnější tvar

$$f(x_j | x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = f(x_j | x_{N_j}). \quad (43)$$



---

**Algorithm 2:** Gibbsův sampler

---

```
zvol  $x^{(0)}$  tak, že  $f(x^{(0)}) > 0$ ;  
for  $i$  from 1 to  $N$  do  
    vygeneruj  $x_1^{(i)} \sim f(x_1|x_2^{(i-1)}, \dots, x_n^{(i-1)})$ ;  
     $\vdots$   
    vygeneruj  $x_j^{(i)} \sim f(x_j|x_1^{(i)}, \dots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \dots, x_n^{(i-1)})$ ;  
     $\vdots$   
    vygeneruj  $x_j^{(i)} \sim f(x_n|x_1^{(i)}, \dots, x_{n-1}^{(i)})$ ;  
end
```

---

Každý krok v každé iteraci Gibbsova sampleru lze chápat jako krok Metropolisova Hastingsova algoritmu. Označme poslední vzorek jako  $x$  a návrh nového vzorku jako  $x^*$ . Je-li v uvažovaném kroku Gibbsova sampleru navrhována nová hodnota  $j$  složky vektoru  $x$ , platí  $x_{-j}^* = x_{-j}$ , kde  $x_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$  a zároveň platí

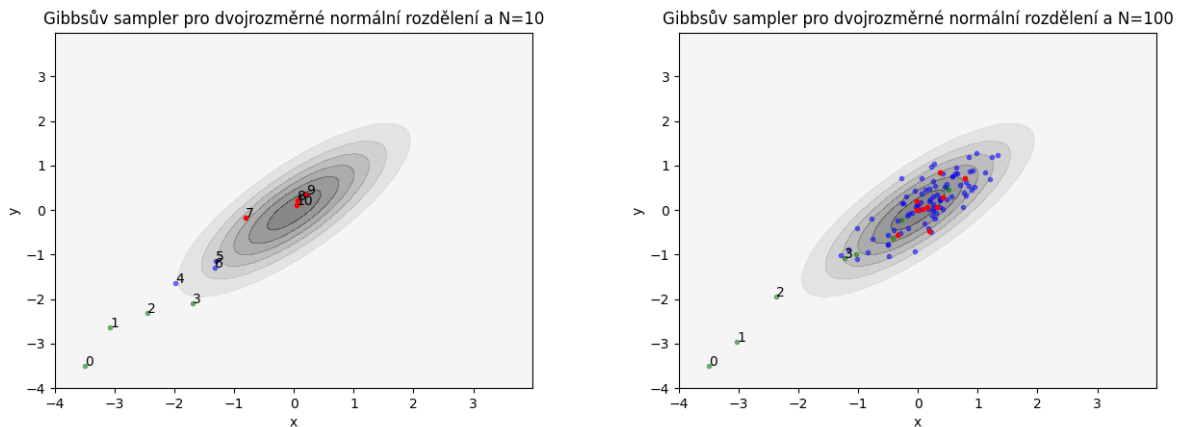
$$\begin{aligned} f(x_j^*|x_{-j}^*) &= f(x_j^*|x_{-j}) \\ q(x_j^*|x) &= f(x_j^*|x_{-j}) \\ q(x_j|x^*) &= q(x_j|x_{-j}^*) = q(x_j|x_{-j}) = f(x_j|x_{-j}). \end{aligned}$$

Pravděpodobnosti přijetí vzorku  $x^*$  je pak rovna

$$\frac{f(x^*)q(x|x^*)}{f(x)q(x^*|x)} = \frac{f(x_j^*|x_{-j})f(x_j|x_{-j})}{f(x_j|x_{-j})f(x_j^*|x_{-j})} = 1,$$

vzorek je tedy přijat vždy (s pravděpodobností 1).

Na obrázku (5) je ukázka provedení Gibbsova sampleru pro generování z dvojrozměrného normálního rozdělení s hustotou  $f(x, y)$ , nulovými středními hodnotami, jednotkovými rozptyly a  $Cov(X, Y) = 0.8$ . Na obrázku vidíme dva samplery se startovní pozicí  $(x^{(0)}, y^{(0)}) = (-3.5, -3.5)$ , vlevo pro  $N = 10$  a vpravo pro  $N = 100$  kroků. Prvních několik vzorků je vyznačeno zeleně, posledních několik naopak červeně.



Obrázek 5: Gibbsův sampler z dvojrozměrného normálního rozdělení pro  $N = 10$  a  $N = 100$

## 6 Teorie míry

V předchozích kapitolách jsme si vysvětlili pojmy náhodné veličiny, vektoru a jejich diskretních a absolutně spojitých rozdělení. My se však s tímto pohledem nespokojíme, budeme sestavovat model, který obsahuje jak absolutně spojitě, tak diskretní náhodné veličiny. S takovými modely si s předchozími definicemi nevystačíme. Aparát teorie míry však ukazuje, že za vhodných podmínek existuje hustota pravděpodobnosti pro obě skupiny náhodných veličin. Při psaní této kapitoly bylo čerpáno z poznámek [6].

**Definice 37.** *Nechť  $\Omega$  je libovolná neprázdná množina  $\mathcal{A} \subset 2^\Omega$ . Řekneme, že  $\mathcal{A}$  je  $\sigma$ -algebra podmnožin  $\Omega$ , platí-li*

- $\Omega \in \mathcal{A}$ ,
- $\forall A \in \mathcal{A} : \bar{A} \in \mathcal{A}$ ,
- $(\forall n \in \mathbb{N} : A_n \in \mathcal{A}) \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$  -  $\sigma$ -aditivita (uzavřenost vůči spočetnému sjednocení).

Uspořádanou dvojici  $(\Omega, \mathcal{A})$  pak nazýváme měřitelným prostorem, prvky množiny  $\mathcal{A}$  měřitelnými množinami.

**Definice 38.** *Nechť  $(\Omega, \mathcal{A})$  je měřitelný prostor. Funkci  $\mu : \mathcal{A} \rightarrow \langle 0, \infty \rangle \cup \{+\infty\}$ , pro kterou platí*

- $\mu(\emptyset) = 0$ ,
- $A_1, A_2, \dots \in \mathcal{A}, \forall i \neq j : A_i \cap A_j = \emptyset \Rightarrow \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ ,

nazveme mírou (na  $\sigma$ -algebře  $\mathcal{A}$ ). Trojici  $(\Omega, \mathcal{A}, \mu)$  pak nazýváme prostorem s mírou. Pokud navíc  $\mu(\Omega) < +\infty$ , míru nazveme konečnou. Jestliže  $\exists M_1, \dots \subset \Omega = \bigcup_{i=1}^{\infty} M_i$  a  $\forall i \in \mathbb{N} : \mu(M_i) < +\infty$ , tak míru nazýváme  $\sigma$ -konečnou.

**Definice 39.** *Nechť  $\Omega$  je neprázdná množina a  $S \subset 2^\Omega$ . Nejmenší  $\sigma$ -algebru obsahující  $S$  nazveme  $\sigma$ -algebrou generovanou množinou  $S$ , značíme  $\sigma S$ .*

**Definice 40.**  $\sigma$ -algebru generovanou systémem všech otevřených množin v  $\mathbb{R}^n$  nazýváme borelovskou  $\sigma$ -algebrou (v  $\mathbb{R}^n$ ), značíme  $\mathcal{B}(\mathbb{R}^n)$ .

**Definice 41.** *Nechť  $\Omega$  je neprázdná množina a  $\mathcal{A} = 2^\Omega$ . Míru  $\mu$  nazveme aritmetickou (čítací), platí-li*

$$\mu(A) = \begin{cases} \text{počet prvků } A & \text{je-li } A \text{ konečná,} \\ +\infty & \text{je-li } A \text{ nekonečná.} \end{cases} \quad (44)$$

**Definice 42.** *Je-li v prostoru  $(\Omega, \mathcal{A}, \mu)$  každá podmnožina množiny nulové míry měřitelná, nazýváme míru  $\mu$  úplnou mírou.*

**Věta 5.** *Na  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  existuje pro  $n \in \mathbb{N}$  právě jedna míra  $\lambda$  taková, že pro každý interval  $I = (a_1, b_1) \times \dots \times (a_n, b_n)$ , kde  $a_n < b_n$  pro  $n \in \mathbb{N}$ , platí  $\lambda(I) = \prod_{i=1}^n (b_i - a_i)$ .*

**Definice 43.** *Míru  $\lambda$  z předchozí věty nazýváme Lebesguovou-Borelovou mírou.*

**Věta 6 (O zúplnění míry).** *Nechť  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou a  $\mathcal{A}_0 = \{E \subset \Omega \mid \exists A, B \in \mathcal{A} : A \subset E \subset B \wedge \mu(B \setminus A) = 0\}$ . Pro  $E \in \mathcal{A}_0$  definujme  $\mu_0(E) = \mu(A)$ . Potom  $(\Omega, \mathcal{A}_0, \mu_0)$  je prostor s úplnou mírou neboli zúplnění  $(\Omega, \mathcal{A}, \mu)$ .*

**Definice 44.** Buď  $(\mathbb{R}^n, \mathcal{B}_0(\mathbb{R}^n), \lambda_0)$  zúplněním  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ . Míru

$$\lambda_0 \tag{45}$$

nazýváme Lebesgueovou mírou a  $\mathcal{B}_0(\mathbb{R}^n)$  systémem lebesgueovsky měřitelných množin.

**Definice 45.** Necht  $(\Omega, \mathcal{A})$  a  $(T, C)$  jsou měřitelné prostory a funkce  $f : \Omega \rightarrow T$ . Řekneme, že funkce  $f$  je měřitelná, jestliže  $\forall A \in C : f^{-1}(A) \in \mathcal{A}$ . Zapisujeme  $f : (\Omega, \mathcal{A}) \rightarrow (T, C)$ . Funkci  $f : (\Omega, \mathcal{A}) \rightarrow (R, \mathcal{B}(\mathbb{R}))$  pak nazýváme borelovsky měřitelnou.

Uvědomme si, že máme-li měřitelné zobrazení  $f : (\Omega, \mathcal{A}) \rightarrow (T, C)$  a míru  $\mu$  na  $(\Omega, \mathcal{A})$ , pak množinová funkce  $\mu_f(B) = \mu(f^{-1}(B))$  je mírou na  $(T, C)$ . Míře  $\mu_f$  na  $(T, C)$  říkáme indukovaná míra.

**Definice 46.** Necht  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou. Uvažujme nějakou vlastnost  $V$  prvků  $\Omega$  (tj.  $V \subset \Omega$ ,  $x \in \Omega$  má vlastnost  $V$ , jestliže  $x \in V$ ). Řekneme, že vlastnost platí pro  $\mu$ -skoro všechna  $x \in \Omega$  ( $\mu$ -skoro všude v  $\Omega$ ), jestliže  $\exists A \in \mathcal{A} : \mu(A) = 0 \wedge (V \text{ platí v } \bar{A})$ .

**Definice 47.** Charakteristická funkce  $\mathbb{I}_A(x) \rightarrow \{0, 1\}$  množiny  $A$  v množině  $\mathcal{A}$  je definována jako

$$\mathbb{I}_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \in \bar{A}. \end{cases}$$

**Definice 48.** Necht  $(\Omega, \mathcal{A})$  je měřitelný prostor. Řekneme, že funkce  $f : \Omega \rightarrow \mathbb{R}$  je jednoduchá funkce, jestliže existují  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$  a  $A_1, \dots, A_n \in \mathcal{A}$  takové, že

$$f = \sum_{i=1}^n \alpha_i \mathbb{I}_{A_i}. \tag{46}$$

Konečně se dostáváme k definici integrálu vzhledem k míře. Ten definujeme postupně pro nezáporné jednoduché funkce, nezáporné funkce a měřitelné funkce. Definujme dále, že  $0 \cdot (+\infty) = 0$ .

**Definice 49.** Necht  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou,  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  je nezáporná jednoduchá funkce a  $\alpha_1, \dots, \alpha_n$  jsou všechny její navzájem různé reálné hodnoty. Označíme-li  $A_i = f^{-1}(\{\alpha_i\})$  pro  $i \in \mathbb{N}$ , pak integrál funkce  $f$  vzhledem k míře  $\mu$ , značíme  $\int f d\mu$  nebo  $\int f(s) d\mu(s)$ , definujeme jako

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i). \tag{47}$$

**Definice 50.** Necht  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou a  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  je nezáporná funkce. Integrál funkce  $f$  vzhledem k míře  $\mu$  definujeme jako

$$\int f d\mu = \sup \left\{ \int g d\mu \mid g : \Omega \rightarrow \mathbb{R} \text{ je jednoduchá funkce } \wedge 0 \leq g \leq f \right\}. \tag{48}$$

**Definice 51.** Necht  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou a  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Označíme-li  $f^+ = \max(f, 0)$  a  $f^- = \max(-f, 0)$ , pak integrál funkce  $f$  vzhledem k míře  $\mu$  definujeme jako

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu, \tag{49}$$

má-li pravá strana smysl. Pokud  $\int f d\mu$  existuje a je konečný, řekneme, že funkce  $f$  je integrovatelná.

**Definice 52.** Necht  $(\Omega, \mathcal{A}, \mu)$  je prostor s mírou,  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  a  $A \in \mathcal{A}$ . Integrál funkce  $f$  přes množinu  $A$  vzhledem k míře  $\mu$  definujeme jako

$$\int_A f d\mu = \int \mathbb{I}_A f d\mu. \quad (50)$$

**Definice 53.** Necht  $(\Omega, \mathcal{A})$  je měřitelný prostor,  $\mu_1, \mu_2$  jsou míry na tomto prostoru a platí  $\forall A \in \mathcal{A} : \mu_1(A) = 0 \Rightarrow \mu_2(A) = 0$ . Pak říkáme, že míra  $\mu_2$  je absolutně spojitá vzhledem k míře  $\mu_1$ , značíme  $\mu_2 \ll \mu_1$ . Míru  $\mu_1$  nazýváme dominující mírou vzhledem k míře  $\mu_2$ .

**Věta 7 (Radonova-Nikodýmova věta).** Necht  $(\Omega, \mathcal{A})$  je měřitelný prostor,  $\mu_1, \mu_2$  jsou míry na tomto prostoru,  $\mu_1$  je  $\sigma$ -konečná a  $\mu_2 \ll \mu_1$ . Pak existuje funkce  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  taková, že  $\forall A \in \mathcal{A} :$

$$\mu_2(A) = \int_A f(x) d\mu_1(x). \quad (51)$$

Funkce  $f$  je jednoznačně určena až na množinu  $\mu_1$  - míry 0, tj.  $\mu_1$ -skoro všude. Je-li funkce  $g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  integrovatelná, pak

$$\int_A g(x) d\mu_2(x) = \int_A g(x)f(x) d\mu_1(x). \quad (52)$$

**Definice 54.** Každou funkci  $f$  z předchozí věty nazýváme Radonovou-Nikodýmovou derivací míry  $\mu_2$  vzhledem k míře  $\mu_1$ , značíme

$$\frac{d\mu_2}{d\mu_1}. \quad (53)$$

Propojíme-li zde uvedené definice a věty se znalostmi teorie pravděpodobnosti, tak jsme schopni říci, že pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$  je vlastně měřitelný prostor s mírou  $P$ , kde  $P(\Omega) = 1$  a prvky  $\sigma$ -algebry  $\mathcal{A}$  jsou náhodné jevy.

Uvažujme tedy pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ . Náhodný vektor je obecně zobrazení

$$X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \quad n \in \mathbb{N},$$

náhodná veličina je speciálním případem, kdy  $n = 1$ .

Uvažujme dále měřitelný prostor  $(\mathcal{X}, \mathcal{B})$  a měřitelné zobrazení  $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$ , pak indukovanou pravděpodobnostní míru  $P_X$  na  $(\mathcal{X}, \mathcal{B})$  indukovanou zobrazením  $X$  z  $P$  nazveme rozdělením pravděpodobnosti měřitelného zobrazení  $X$ , tedy náhodné veličiny nebo vektoru. Pravděpodobnostní prostor se tedy náhodnou veličinou transformuje na prostor  $(\mathcal{X}, \mathcal{B}, P_X)$ . Předpokládejme dále, že indukovaná míra  $P_X$  je absolutně spojitá vzhledem k  $\nu$  na  $(\mathcal{X}, \mathcal{B})$ . Zvolme  $B \in \mathcal{B}$  a za měřitelnou funkci  $g : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  zvolme charakteristickou funkci  $\mathbb{I}_B$ . Radonovu-Nikodýmovu větu (52) tak můžeme vyjádřit jako

$$\int_{\mathcal{X}} \mathbb{I}_B dP_X(x) = \int_B f_X(x) d\nu(x). \quad (54)$$

Všimněme si, že  $f_X$ , tedy Radonova-Nikodýmova derivace  $\frac{dP_X}{d\nu}$ , je hustotou pravděpodobnosti měřitelného zobrazení  $X$  vzhledem k míře  $\nu$ .

Na základě těchto myšlenek jsme schopni říci, zvolíme-li vhodnou míru  $\nu$ , která bude  $\sigma$ -konečná a  $P_X$  bude absolutně spojitá vzhledem k  $\nu$ , že náhodná veličina (náhodný vektor)  $X$  má hustotu pravděpodobnosti. Uvažujme teď čítací (44) a Lebesgueovu míru (45). Má-li náhodná veličina  $X$  diskrétní rozdělení, tak  $P_X$  je absolutně spojitá vzhledem k čítací míře, vzhledem ke které pak existuje také hustota pravděpodobnosti  $f_X$  náhodné veličiny  $X$ . Má-li rozdělení absolutně spojité, pak je  $P_X$  absolutně spojitá vzhledem k míře Lebesgueově, vzhledem ke které pak hustota  $f_X$  také existuje.

## 7 Bayesovská rozhodovací úloha

Tato kapitola byla sepsána na základě [7] a inspirací byl, včetně značení, text [8]. Předpokládejme, že náhodné veličiny mají hustotu pravděpodobnosti vzhledem k čítací míře (44) v případě diskrétní náhodné veličiny a k míře Lebesgueově (45) v případě absolutně spojitě náhodné veličiny.

Bayesovské metody jsou vhodným nástrojem pro řešení problémů, ve kterých jsme schopni ke zkoumaným datům před jejich pozorováním přidat nějakou informaci, tzv. apriorní informaci, která následně zpřesní řešení úlohy. Princip bayesovské rozhodovací úlohy tkví v tom, že všechny neznámé parametry jsou považovány za náhodné. Neznámé náhodné parametry vstupují do úlohy ve formě parametru  $\theta$ , na kterém závisí rozdělení zkoumané náhodné veličiny  $x$ .

Jako statistický model rozhodovací úlohy chápeme systém hustot pravděpodobností reprezentovaný podmíněnou hustotou  $x$  za podmínky parametru  $\theta$

$$f(x|\theta). \quad (55)$$

Tuto hustotu jako funkci parametru  $\theta$  nazýváme věrohodnostní funkcí.

Apriorní hustotou pravděpodobnosti nazveme hustotu parametru  $\theta$

$$f(\theta), \quad (56)$$

která do rozhodovací úlohy přináší již zmíněnou informaci před pozorováním dat. Při volbě apriorní hustoty se často vychází ze zkušenosti či fyzikálních modelů, podléhá tedy subjektivitě řešitele dané úlohy.

Klíčovou roli v bayesovské rozhodovací úloze hraje tzv. aposteriorní hustota pravděpodobnosti  $f(\theta|x)$ , kterou lze z Bayesovy věty (7) vyjádřit jako

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta) d\theta}. \quad (57)$$

Aposteriorní hustota tedy na základě apriorní hustoty a pozorování dat představuje souhrnnou informaci o neznámém parametru  $\theta$ . Hustota (57) je plně určena čitatelem,  $f(x)$  je konstanta vzhledem k  $\theta$ . Můžeme ji pak napsat ve tvaru

$$f(\theta|x) \propto f(x|\theta)f(\theta). \quad (58)$$

Symbol  $\propto$  vyjadřuje úměrnost levé strany ku pravé.  $f(x)$  často kvůli počtu dimenzí a s tím spojené výpočetní náročnosti určit, proto se pak užívá vztahu (58). Někdy však ani tato simplifikace nepomůže a pro vyřešení integrálů v aposteriorních hustotách nestačí ani klasické numerické metody, v takových případech se používají simulační metody, například metody MCMC (Markov chain Monte Carlo).

Bayesovská rozhodovací úloha je založena na uskutečnění nějakého rozhodnutí. Všechna rozhodnutí jsou sledována pomocí tzv. ztrátové funkce

$$L : \mathcal{A} \times \Theta \rightarrow \langle 0, +\infty \rangle, \quad (59)$$

kde  $\mathcal{A}$  je množina všech rozhodnutí a  $\Theta$  je obor hodnot  $\theta$ .

Ztrátová funkce přiřazuje každému rozhodnutí  $a \in \mathcal{A}$  v závislosti na hodnotě parametru  $\theta$  reálné číslo popisující ztrátu při volbě daného rozhodnutí. Z tohoto vyplývá, že minimalizace střední hodnoty ztrátové funkce povede k optimálnímu rozhodnutí  $a^{opt}$ , tedy

$$a^{opt} \in \underset{a \in \mathcal{A}}{\text{Argmin}} \int L(a, \theta) f(x|\theta) d\theta, \quad (60)$$

kde Argmin je množina argumentů minima.

**Příklad 3.** Klasickým příkladem je odvození aposteriorní hustoty parametru střední hodnoty normálního rozdělení. Předpokládejme tedy, že  $x_1, \dots, x_t$  jsou nezávislé, stejně rozdělené náhodné veličiny s normálním rozdělením

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

s neznámým náhodným parametrem  $\mu$  a známým rozptylem  $\sigma^2$ . Protože jsou veličiny  $x_1, \dots, x_n$  nezávislé, můžeme psát

$$f(x_1, \dots, x_t|\mu) = \prod_{\tau=1}^t f(x_\tau|\mu) = -\frac{1}{\sqrt{(2\pi\sigma^2)^t}} \exp\left(-\frac{1}{2\sigma^2} \sum_{\tau=1}^t (x_\tau - \mu)^2\right).$$

Předpokládejme, že apriorní informace o parametru  $\mu$  lze vyjádřit pomocí normálního rozdělení, a to s parametry  $(\mu_0, \sigma_0^2)$ . Hustota

$$f(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right)$$

je tedy apriorní hustotou.

A protože jsou  $x_1, \dots, x_t$  podmíněně nezávislé za podmínky  $\mu$ , můžeme psát vztah pro aposteriorní hustotu (57)

$$\begin{aligned} f(\mu|x_1, \dots, x_t) &\propto f(x_1, \dots, x_t|\mu) f(\mu) \\ &\propto (2\pi\sigma^2)^{-\frac{t}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{\tau=1}^t (x_\tau - \mu)^2\right) (2\pi\sigma_0^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{\tau=1}^t (x_\tau - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right) \\ &= \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{\tau=1}^t x_\tau^2 + 2\mu \sum_{\tau=1}^t x_\tau + t\mu^2\right) - \frac{1}{2\sigma_0^2} (\mu^2 - 2\mu_0\mu + \mu_0^2)\right) \\ &= \exp\left(-\frac{\mu^2}{2} \left(\frac{1}{\sigma_0^2} + \frac{t}{\sigma^2}\right) + \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{\tau=1}^t x_\tau}{\sigma^2}\right) - \frac{\mu_0^2}{2\sigma_0^2} - \frac{\sum_{\tau=1}^t x_\tau^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{\mu^2}{2} \left(\frac{1}{\sigma_0^2} + \frac{t}{\sigma^2}\right) + \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{\tau=1}^t x_\tau}{\sigma^2}\right)\right). \end{aligned}$$

Označíme-li si pak

$$\sigma_1^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{t}{\sigma^2}}, \quad \mu_1 = \sigma_1^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{\tau=1}^t x_\tau}{\sigma^2}\right),$$

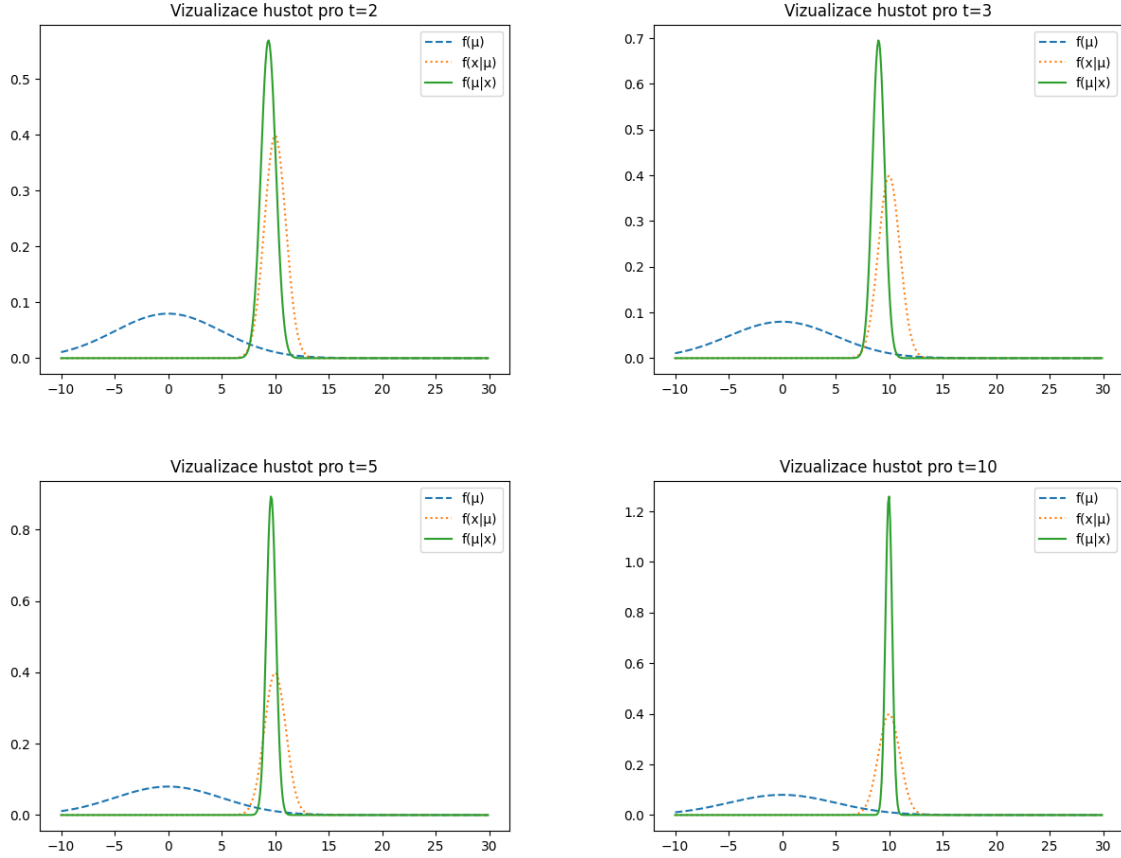
získáváme aposteriorní hustotu ve tvaru

$$f(\mu|x_1, \dots, x_t) \propto \exp\left(-\frac{(\mu - \mu_1)^2}{2\sigma_1^2}\right).$$

Povšimněme si, že aposteriorní rozdělení má pro libovolné přípustné parametry apriorní informace rovněž normální rozdělení  $\mu|x_1, \dots, x_t \sim N(\mu_1, \sigma_1^2)$ .<sup>4</sup>

Správně by teď následovalo určení množiny rozhodnutí, sestavení ztrátové funkce a minimalizace střední hodnoty ztrátové funkce vzhledem k aposteriorní hustotě pro získání optimálního rozhodnutí. Nám pro ilustraci však postačí ukázat tvar aposteriorní hustoty vzhledem k apriorní hustotě a věrohodnostní funkci. Jako parametry apriorní hustoty zvolme  $(\mu_0 = 0, \sigma_0^2 = 25)$ , pro model s pevným rozptylem pak zvolme nějaké parametry  $\mu, \sigma^2$ , například  $(\mu = 10, \sigma^2 = 1)$ .

Ukažme si na obrázku (6) tvary hustot pro různé počty pozorování  $x_1, \dots, x_t$ .



Obrázek 6: Znázornění tvaru aposteriorní hustoty pro rozdílný počet pozorovaných dat

Z obrázku (6) lze vyčíst, že s narůstajícím počtem pozorování se rozložení aposteriorní hustoty postupně koncentruje kolem skutečné hodnoty parametru, což je v souladu s intuicí. Při nedostatku dat hraje velkou roli apriorní informace, ta se s přibývajícím evidencí tolik na tvaru aposteriorní hustoty neprojevuje. Zároveň vidíme, že aposteriorní rozptyl s přibývajícím daty klesá, což se projeví na tvaru rozdělení, bude totiž méně variabilní, tedy více koncentrované u střední hodnoty.

<sup>4</sup>Množinu (rodinu)  $\mathcal{F}$  hustot parametru  $\theta$  nazýváme konjugovanou rodinou hustot parametru  $\theta$  pro model  $f(x|\theta)$ , jestliže pro každé apriorní rozdělení  $f(\theta) \in \mathcal{F}$  a každé pozorování  $x$  je aposteriorní rozdělení  $f(\theta|x)$  rovněž z  $\mathcal{F}$ .

## 8 Segmentace CT snímku aorty břišní

CT (computed tomography) je primárně medicínská zobrazovací metoda, která za použití rentgenových paprsků umožňuje zobrazit zkoumanou oblast lidského těla (případně tělo celé) v sérii řezů. V našem případě se jedná o snímky aorty břišní pacientů s výdutí aorty. Mnohdy se stává, že je na těchto snímcích špatně rozlišitelné, o jakou část zkoumané oblasti se jedná. V případě snímků aorty se jedná primárně o prostor, kterým protéká krev (lumen), stěnu cévy s krevní sraženinou (trombus) a tkáň jakožto vnější prostor aorty. Předpokládejme tedy, že obraz tvoří tři souvislé plochy s odlišnou náhodnou texturou.

CT snímek je obdélníkový rastrový obrázek v odstínech šedi, což znamená, že je tvořen jednotlivými pixely tak, že barvu (jas) každého z  $n \times m$  pixelů lze vyjádřit hodnotou z množiny  $\{0, \dots, 255\}$ . Z předpokladu tedy plyne, že pro každý pixel z nějaké ze tří zmiňovaných oblastí bude pravděpodobnější, že se vedle něj nachází pixely ze stejné oblasti. Na základě těchto myšlenek problém přirozeně vede k bayesovské rozhodovací úloze, kdy sestavíme vhodný model a s pomocí apriorní informace (celistvost oblastí) a informací z dat (pozorované hodnoty jasu) dokážeme vhodně odhadnout segmentaci snímku, tedy rozdělení snímku na lumen, trombus a okolí aorty. Proč má segmentace CT snímků pacientů s výdutí aorty břišní takový význam se můžeme dočíst v článku [9]. Pro formulaci problému je využito textu [8].

Protože obraz tvoří  $n \times m$  pixelů, bylo by vhodné si je nějakým způsobem označit. Každý pixel tedy můžeme vhodně indexovat použitím množiny  $S$  jakožto množiny uspořádaných dvojic takové, že

$$S = \{1, \dots, n\} \times \{1, \dots, m\}.$$

Každá dvojice  $(i, j) \in S$  (chápejme jako souřadnice) tedy označuje jeden daný pixel.

Uvažujme nyní náhodná pole

$$x = (x_{ij})_{(i,j) \in S},$$

$$y = (y_{ij})_{(i,j) \in S}.$$

Pro každý pixel  $(i, j) \in S$  představuje náhodné pole  $x$  hodnotu segmentace v daném bodě. Jelikož segmentujeme obraz do tří oblastí, bude  $x_{ij}$  nabývat vždy jedné z hodnot z množiny  $Q = \{0, 1, 2\}$ . Opravdové hodnoty náhodného vektoru  $x$  nejsou známy, vstupuje tedy do úlohy v roli parametru.

Pole  $y$  pak v každém pixelu popisuje pozorovanou hodnotu jasu. Na každý pixel jsou tedy vázány dvě náhodné veličiny,  $y_{ij}$  popisující opravdovou hodnotu jasu a  $x_{ij}$  popisující hodnotu segmentace. Předpokládejme pro jednoduchost, že pozorované veličiny  $y_{ij}$  mají normální rozdělení se známými parametry  $(\mu_q, \sigma_q^2)$  pro  $q \in Q$  v závislosti na oblasti, ve které se nachází. Pro každý pixel  $(i, j) \in S$  dostáváme model

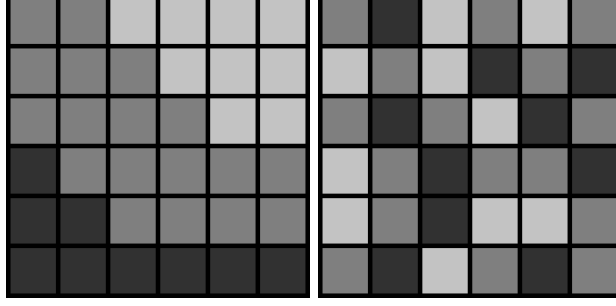
$$f(y_{ij}|x_{ij} = q) = \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2}\right). \quad (61)$$

A jelikož víme, že pozorované hodnoty  $y_{ij}$  jsou podmíněně nezávislé, když známe  $x$ , a podmíněně nezávislé na hodnotách  $x_{kl}$  jiných pixelů než  $(i, j)$ , tak můžeme model vyjádřit jako

$$f(y|x) = \prod_{(i,j) \in S} f(y_{ij}|x_{ij}). \quad (62)$$

Pro sestavení aposteriorní hustoty potřebujeme ještě zvolit apriorní hustotu  $f(x)$ . Uvědomme si, že předpokládáme celistvost segmentovaných oblastí, tuto skutečnost bychom tedy v této hustotě mohli nějakým způsobem reflektovat. V úlohách zpracování obrazu se často používají Gibbsovy





Obrázek 7: Znázornění segmentací obrázku  $6 \times 6$  vlevo s menší energií, vpravo s větší energií

distribuce (37). Jedná se o pravděpodobnostní rozdělení, která ve statistické mechanice popisují pravděpodobnostní rozdělení stavů velkého počtu částic. Tato rozdělení mají tvar (viz. 4.3)

$$P(X = x) = \frac{\exp(-\beta H(x))}{\sum_{\tilde{x} \in \mathcal{X}} \exp(-\beta H(\tilde{x}))}. \quad (63)$$

Funkce  $H(x)$  se interpretuje jako energie systému a v Gibbsových distribucích hraje klíčovou roli. Ve zpracování obrazu se používají modely, kdy je funkce energie  $H(x)$  vyjádřena jako součet příspěvků vzhledem k nějakým malým skupinám pixelů. Konstanta  $\beta$  určuje nakolik se změna energie projeví na změně hustoty. Pro množinu  $S$  tedy navrhneme systém okolí  $\mathcal{N}$  (viz 4.2) tak, že každý pixel  $(i, j) \in S$  má za množinu sousedů pixely přímo pod, nad, vpravo a vlevo od něj. Označme si množinu sousedů pixelu  $(i, j)$  jako  $\mathcal{N}(i, j)$ . Pak platí

$$\mathcal{N}(i, j) = \{(k, l) \in S : |i - k| + |j - l| = 1\}. \quad (64)$$

Energii  $H(x)$  můžeme pro danou segmentaci (náhodné pole  $x$ ) vyjádřit jako součet počtů všech sousedních pixelů s různou hodnotou segmentace přes všechny pixely

$$H(x) = \sum_{(i,j) \in S} \left( \sum_{(k,l) \in \mathcal{N}(i,j)} 1 - \delta(x_{ij}, x_{kl}) \right), \quad (65)$$

kde  $\delta(a, b)$  (Kroneckerovo  $\delta$ ) je funkce

$$\delta(a, b) = \begin{cases} 1 & \text{když } a = b, \\ 0 & \text{když } a \neq b. \end{cases} \quad (66)$$

Hodnota  $H(x)$  je tedy přímo úměrná délce hranice mezi oblastmi s různými hodnotami segmentace. Uvědomme si, že hustotu (61) můžeme upravit do tvaru vhodnějšího pro výpočet

$$\begin{aligned} f(y_{ij}|x_{ij} = q) &= \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2}\right), \\ f(y_{ij}|x_{ij}) &= \prod_{q \in Q} \left( \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2}\right) \right)^{\delta(x_{ij}, q)}. \end{aligned} \quad (67)$$

Nyní, když známe tvar apriorní hustoty (63) a modelu (67), můžeme sestavit aposteriorní hustotu

$$\begin{aligned}
f(x|y) &= \frac{f(y|x)f(x)}{f(y)} \\
&\propto \left( \prod_{(i,j) \in S} f(y_{ij}|x_{ij}) \right) f(x) \\
&= \left( \prod_{(i,j) \in S} \prod_{q \in Q} \left( \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp \left( -\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2} \right) \right)^{\delta(x_{ij},q)} \right) \frac{\exp(-H(x))}{\sum_{x \in \mathcal{X}} \exp(-H(x))} \\
&\propto \left( \prod_{(i,j) \in S} \prod_{q \in Q} \left( \frac{1}{\sigma_q^2} \exp \left( -\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2} \right) \right)^{\delta(x_{ij},q)} \right) \exp(-H(x)).
\end{aligned} \tag{68}$$

Pro uskutečnění optimálního rozhodnutí nalezneme bodový odhad  $\hat{x}$  náhodného pole  $x$  představující segmentaci snímku. Množina všech rozhodnutí jsou všechna rozložení tříbarevných pixelů na obrázku  $n \times m$ , tedy  $\{0, 1, 2\}^{(n \times m)}$ . Naším cílem je vybrat to nejlepší, tedy to, které minimalizuje střední hodnotu ztrátové funkce vzhledem k aposteriornímu rozdělení

$$\hat{x} \in \underset{a \in Q^{(n \cdot m)}}{\operatorname{Argmin}} E[L(a, x)|y]. \tag{69}$$

Tvar aposteriorní hustoty už známe, zbývá nám zvolit vhodnou ztrátovou funkci. V oblasti zpracování obrazu se používá tzv. 0-1 ztrátová funkce ve tvaru

$$L(a, x) = 1 - \delta(a, x) = \begin{cases} 1 & a \neq x, \\ 0 & a = x. \end{cases} \tag{70}$$

Střední hodnotu 0-1 ztrátové funkce vzhledem k aposteriornímu rozdělení můžeme upravit do tvaru

$$\begin{aligned}
E[L(a, x)|y] &= \int L(a, x) f(x|y) dx \\
&= \int (1 - \delta(a, x)) f(x|y) dx \\
&= \int f(x|y) dx - \int \delta(a, x) f(x|y) dx \\
&= 1 - f(x = a|y),
\end{aligned} \tag{71}$$

ze kterého vyplývá, že pro získání minimální očekávané ztráty musíme maximalizovat aposteriorní hustotu

$$\hat{x} \in \underset{a \in Q^{(n \times m)}}{\operatorname{Argmax}} f(a|y). \tag{72}$$

Pro nalezení maxima aposteriorní hustoty (68) se kvůli dimenzi úlohy neobejdeme bez simulačních metod typu Markov Chain Monte Carlo. Tyto metody slouží primárně pro numerickou integraci, zároveň je však lze použít pro nalezení přibližného řešení optimalizační úlohy (72). Aposteriorní rozdělení v tomto typu úloh bývá velmi koncentrované - „typický“ vzorek z aposteriorní hustoty můžeme často považovat za dostatečně dobrou aproximaci. Vzhledem k snadnému nalezení potřebných podmíněných hustot se pro naši úlohu nabízí tzv. Gibbsův sampler (2) popsáný v sekci (5.2). V úlohách, kde složky náhodného vektoru závisí jen na některých z ostatních složek, se generování takových vzorků značně zjednodušuje.

Pro každý pixel  $(i, j) \in S$  platí

$$f(x_{ij}|x_{kl_{(k,l) \in S \setminus \{(i,j)\}}}, y) = f(x_{ij}|x_{kl_{(k,l) \in \mathcal{N}(i,j)}}, y), \quad (73)$$

což znamená, že aposteriorní hustota každého pixelu za podmínky všech ostatních je rovna aposteriorní hustotě toho pixelu za podmínky sousedů.

Postupně generujeme náhodné pole  $x$ , díky znalosti (73) generujeme vzorek z hustoty

$$x_{ij} \sim f(x_{ij}|x_{kl_{(k,l) \in \mathcal{N}(i,j)}}, y). \quad (74)$$

Předchozí vztah můžeme upravit do tvaru

$$f(x_{ij}|x_{kl_{(k,l) \in \mathcal{N}(i,j)}}, y) = \frac{f(x_{ij}, x_{kl_{(k,l) \in \mathcal{N}(i,j)}}|y)}{f(x_{kl_{(k,l) \in \mathcal{N}(i,j)}}|y)} = \frac{f(x_{ij}, x_{kl_{(k,l) \in \mathcal{N}(i,j)}}|y)}{\sum_{q \in Q} f(x_{ij} = q, x_{kl_{(k,l) \in \mathcal{N}(i,j)}}|y)}. \quad (75)$$

Na základě předpisu pro aposteriorní hustotu (68) a úvahy, že přírůstek energie  $H(x)$  nezávisí na bodech mimo okolí uvažované oblasti, jsme schopni tvrdit, že

$$\begin{aligned} f(x_{ij}|x_{kl_{(k,l) \in \mathcal{N}(i,j)}}, y) &\propto \left( \prod_{(\bar{i}, \bar{j}) \in \mathcal{N}(i,j) \cup \{(i,j)\}} \prod_{q \in Q} \left( \frac{1}{\sigma_q^2} \exp \left( -\frac{(y_{\bar{i}\bar{j}} - \mu_q)^2}{2\sigma_q^2} \right) \right)^{\delta(x_{\bar{i}\bar{j}}, q)} \right) \\ &\times \exp \left( -\beta \left( \sum_{(\bar{i}, \bar{j}) \in \mathcal{N}(i,j) \cup \{(i,j)\}} \left( \sum_{(k,l) \in \mathcal{N}(\bar{i}, \bar{j})} 1 - \delta(x_{\bar{i}\bar{j}}, x_{kl}) \right) \right) \right), \end{aligned} \quad (76)$$

což můžeme dle vztahu (75) upravit do tvaru.

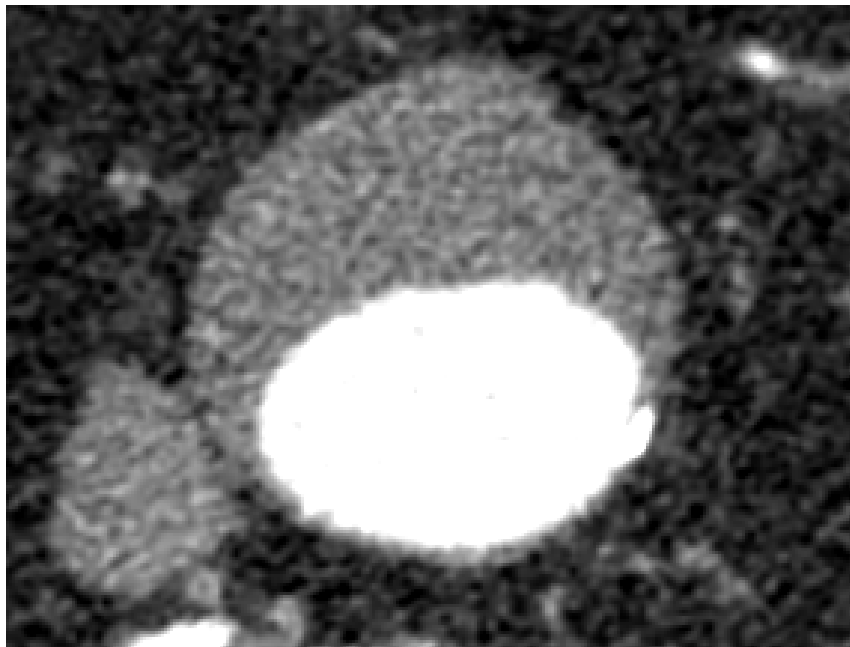
$$\propto \prod_{q \in Q} \left( \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp \left( -\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2} \right) \right)^{\delta(x_{ij}, q)} \exp \left( -\beta \left( \sum_{(k,l) \in \mathcal{N}(i,j)} 1 - \delta(x_{ij}, x_{kl}) \right) \right) \quad (77)$$

Vztah určuje hustotu  $f(x_{ij}|x_{kl_{(k,l) \in \mathcal{N}(i,j)}}, y)$  až na normalizační konstantu rovnu

$$\left( \sum_{\tilde{x}_{ij} \in Q} \prod_{q \in Q} \left( \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp \left( -\frac{(y_{ij} - \mu_q)^2}{2\sigma_q^2} \right) \right)^{\delta(\tilde{x}_{ij}, q)} \exp \left( -\beta \left( \sum_{(k,l) \in \mathcal{N}(i,j)} 1 - \delta(\tilde{x}_{ij}, x_{kl}) \right) \right) \right)^{-1}.$$

Nyní máme veškeré potřebné informace k úspěšnému běhu Gibbsova sampleru. Generujeme tedy postupně vzorky  $x_{11}^{(1)}, x_{12}^{(1)}, \dots, x_{nm}^{(1)}, x_{11}^{(2)}, x_{12}^{(2)}, \dots, x_{nm}^{(2)}, \dots, x_{11}^{(\iota)}, x_{12}^{(\iota)}, \dots, x_{nm}^{(\iota)}$  z hustot ze vztahu (77) až do dosáhnutí počtu zvolených iterací  $\iota$ . Jak již bylo zmíněno, díky koncentraci aposteriorní hustoty budou „typické“ vzorky dostatečnou aproximací segmentace.

Ukažme si nyní původní CT snímek, tedy jinými slovy náhodné pole  $y$ , na kterém budeme provádět proces segmentace.



Obrázek 8: CT snímek aorty břišní

Na snímku (8) jsou vyobrazeny všechny tři části, které potřebujeme rozeznat. Bílá oblast je prostor, kterým teče krev (lumen), šedá oblast je sraženina (trombus) a nejtmaší oblast je prostor mimo cévu. Snímek je zanesen šumem, mnohdy proto nelze přesně určit, do které oblasti zvolený pixel patří. Na základě toho je zřejmé, proč má pokus o určení nejpravděpodobnějšího rozložení oblastí na CT snímku takový význam.

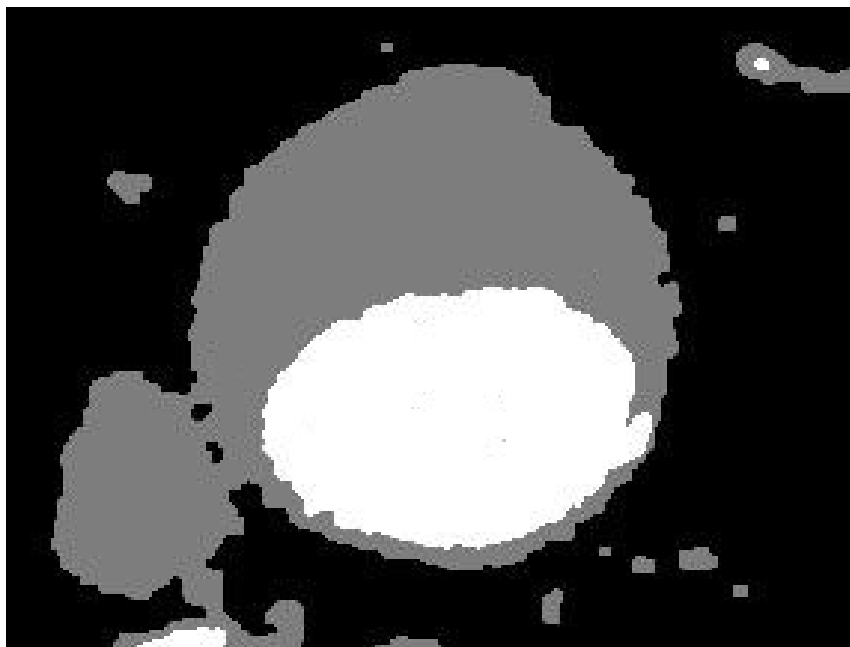
Pro finální nalezení hodnot pole  $x$  zbývá už pouze vhodně zvolit parametry modelu  $(\mu_0, \sigma_0^2)$ ,  $(\mu_1, \sigma_1^2)$ ,  $(\mu_2, \sigma_2^2)$  a konstantu  $\beta$  v apriorním rozdělení (63). Odhad těchto parametrů získáme jako výběrový průměr a rozptyl z předem vybraných vzorových oblastí z dané části obrázku. Zvolme tedy, že

$$x_{ij} = \begin{cases} 0, & \text{je-li v lumenu,} \\ 1, & \text{je-li v trombu,} \\ 2, & \text{je-li ve vnějším prostoru.} \end{cases} \quad (78)$$

Na základě nalezených odhadů zvolme parametry  $(\mu_0, \sigma_0^2) = (40, 640)$ ,  $(\mu_1, \sigma_1^2) = (124, 256)$  a  $(\mu_2, \sigma_2^2) = (255, 1)$ . Z hodnot získaných výběrových charakteristik si lze povšimnout jisté podobnosti s obrázkem (8). Všimněme si, že lumen je jasně bílý s minimem odlehých hodnot, čemuž také odpovídají hodnoty  $\mu_2 = 255$  (hodnota bílé barvy v barevném spektru) a  $\sigma_2^2 = 1$  (velmi malý rozptyl). Naopak v oblasti mimo aortu vidíme větší množství světlejších oblastí, rozptyl  $\sigma_0^2 = 640$  této oblasti je tedy řádově větší.

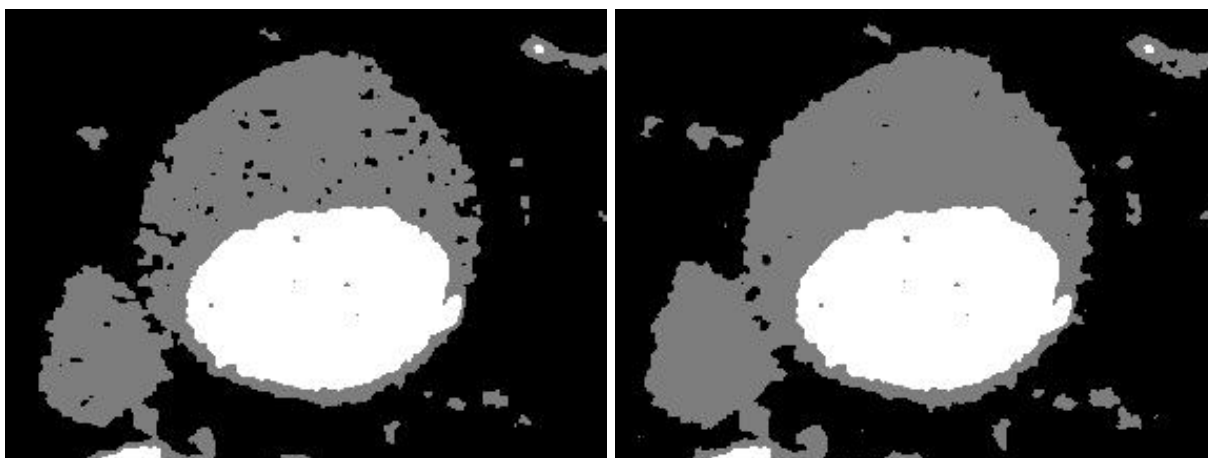
Zvýšení konstanty  $\beta$  přidává váhu apriorní informaci. Zvolme tedy  $\beta = 2$ , zvolme  $f(x^{(0)})$  jako nulové pole a nechme Gibbsův sampler běžet 15 iterací. Podívejme se na výsledek na obrázku (9), kde jsou hodnoty  $x_{ij} = 0$  vyznačeny bíle,  $x_{ij} = 1$  šedě a  $x_{ij} = 2$  černě.

Vidíme, že už pro 15 iterací je výsledek velice uspokojivý. Snímek je, až na výjimky (nejspíš přilehlé menší cévy, které tvoří malou souvislou oblast mimo aortu), tvořen třemi velkými souvislými oblastmi jasně danými jednou ze tří barev. Uvnitř trombu i lumenu pak nevidíme žádné známky oblastí, které tam zajisté nepatří.



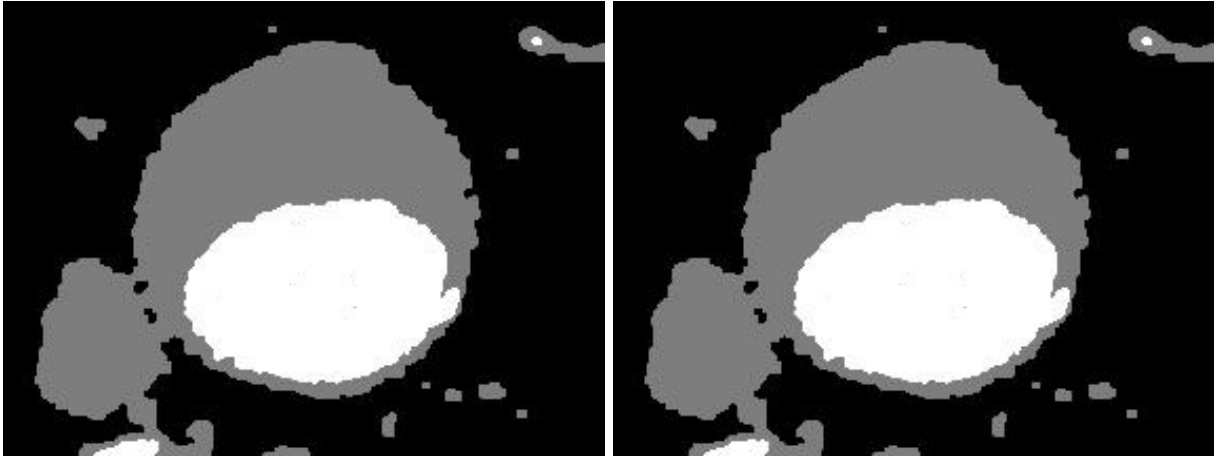
Obrázek 9: Výsledná segmentace s vhodnými parametry po 15 iteracích

Na obrázku (10) pak sledujme vývoj algoritmu pro 1 a 5 iterací s původními parametry. Je vidět, že už po první iteraci Gibbsova sampleru vzniká velice přívětivý obrázek, značně podobný výsledné segmentaci. Jak již bylo několikrát zmiňováno, v každé další iteraci se projde celý obrázek a dělá se korektura a aktualizace příslušných pixelů.



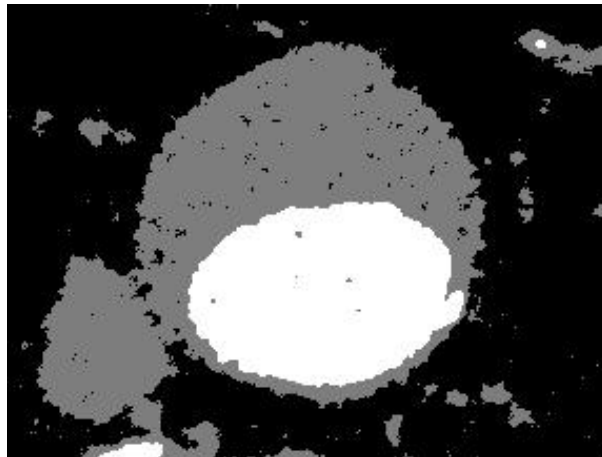
Obrázek 10: Vlevo segmentace po 1 a vpravo po 5 iteracích

Ukažme si dále na obrázku (11), že Gibbsův sampler jakožto speciální případ Metropolisova-Hastingsova algoritmu opravdu dosáhne stacionárního rozdělení (jinými slovy se obrázky od určité iterace již budou jen velice málo měnit, což ilustruje fakt, že aposteriorní rozdělení je velmi koncentrované a typické vzorky lze považovat za vhodnou aproximaci optimální segmentace). Vlevo máme 15 a vpravo 50 iterací Gibbsova sampleru, vidíme, že obrázky jsou téměř totožné, generovaný markovský řetězec dosáhl stacionárního rozdělení.



Obrázek 11: Ukázka stacionárního rozdělení Gibbsova sampleru

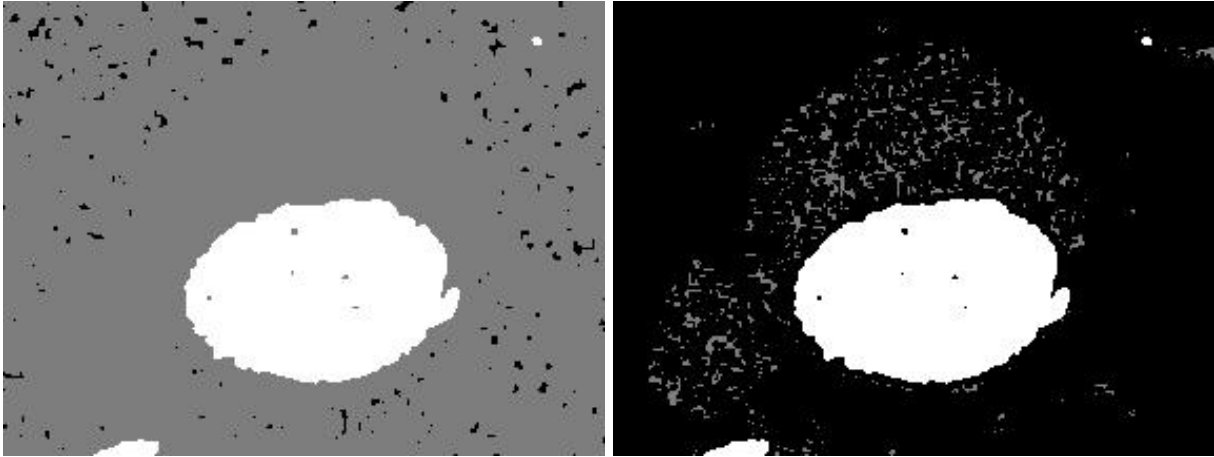
Dále by bylo vhodné si ukázat, jak by segmentace vypadala, kdybychom zvolili odlišné parametry. První si zkusme ukázat, jak by vypadala segmentace, kdybychom apriorní informaci dali menší váhu, tedy zmenšili konstantu  $\beta$ . Na obrázku (12) jsme nechali stejné střední hodnoty a rozptyly, ale zvolili  $\beta = \frac{1}{10}$ . Opět jsme algoritmus nechali běžet 15 iterací.



Obrázek 12: Segmentace po 15 iteracích s  $\beta = \frac{1}{10}$ .

Vidíme, že na rozdíl od obrázku (9), kde jsme dávali apriorní informaci větší váhu, v tomto případě algoritmus více věří informacím reálných hodnot jasu CT snímku (tedy hodnotám pole  $y$ ), na obrázku se tedy nachází znatelně více diskontinuit v jednotlivých oblastech. Je zřejmé, že tohle by správná volba nebyla, uvnitř trombu se jen stěží bude nacházet vnějšek aorty apod.

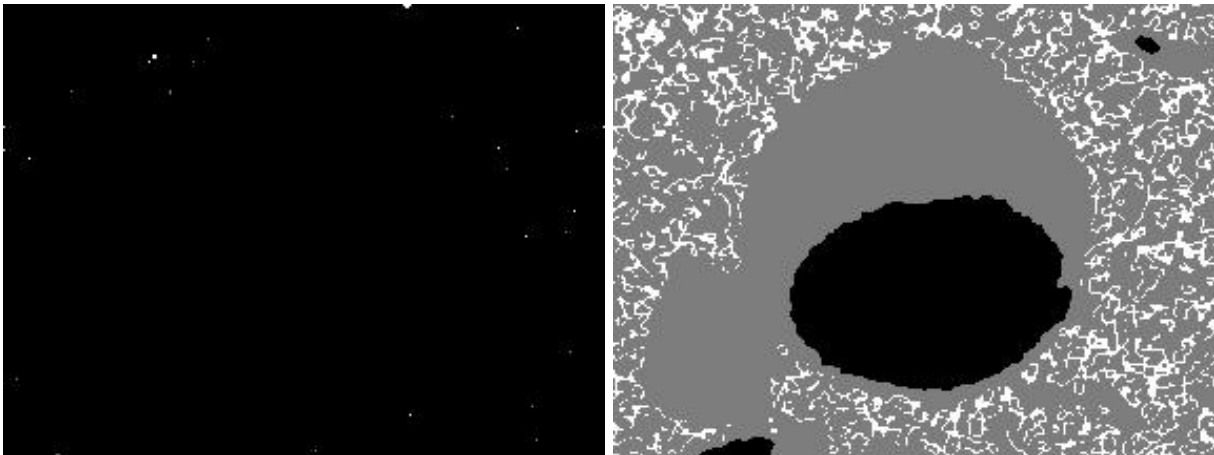
Nakonec si ještě ukažme několik případů, kdy byly nevhodně zvoleny i ostatní parametry, tedy střední hodnoty a rozptyly modelu. Na obrázku (13) vidíme dvě různé segmentace Gibbsova sampleru s 15 iteracemi s odlišnými parametry.



Obrázek 13: Segmentace se změněnými rozptyly

Vlevo je konfigurace se středními hodnotami z původní segmentace a u rozptylů jsme pouze mnohokrát zvýšili rozptyl šedé složky. Parametry tohoto obrázku jsou tedy  $(\mu_0, \sigma_0^2) = (40, 1)$ ,  $(\mu_1, \sigma_1^2) = (124, 2500)$  a  $(\mu_2, \sigma_2^2) = (255, 1)$ . Nastavením rozptylu šedé složky na tak vysokou hodnotu jsme modelu předali informaci, že chceme hodnotu  $x_{ij} = 1$  přiřazovat oblastem s vysokou variabilitou intenzity pixelů. Proto je většina pixelů vnějšího prostoru šedá, přesto, že oblastně leží ve vnějšku aorty, podle parametrů modelu měla blíže k hodnotě 1, tedy stěně aorty, a navíc apriorní informace udává, že se mají spíše tvořit souvislé plochy. Lumen se změny parametru nedotkl, proto zůstává relativně homogenně bílá.

Naopak u obrázku vpravo jsme nechali původní střední hodnoty i rozptyly, až na rozptyl šedé složky, ten jsme tentokrát zmenšili. Parametry jsou tedy  $(\mu_0, \sigma_0^2) = (40, 640)$ ,  $(\mu_1, \sigma_1^2) = (124, 1)$  a  $(\mu_2, \sigma_2^2) = (255, 1)$ . V tomhle případě jsme variabilitu šedých pixelů značně snížili, tím pádem se šedá přiřazuje těm pixelům, které jsou reálně šedé opravdu blízko. Proto je v oblasti trombu velké množství černých pixelů. Bílé složky se změna opět nedotkla.



Obrázek 14: Segmentace se změněnými středními hodnotami

Podívejme se nyní na obrázek (14). V případě vlevo vidíme segmentaci, kde byly nastaveny všechny střední hodnoty na hodnotu černé oblasti a rozptyly zůstaly původní, tedy  $(\mu_0, \sigma_0^2) = (40, 640)$ ,  $(\mu_1, \sigma_1^2) = (40, 256)$  a  $(\mu_2, \sigma_2^2) = (40, 1)$ . Při této volbě jsou rychle zmenšovány

hodnoty hustot v závislosti na vzdálenosti  $y_{ij}$  od střední hodnoty, což nakonec vede k faktu, že jsou s největší pravděpodobností vybírány černé hodnoty, tedy  $x_{ij} = 0$ . V posledním z obrázků jsou přehozeny střední hodnoty, přičemž rozptyly zůstávají stejné, tedy  $(\mu_0, \sigma_0^2) = (255, 640)$ ,  $(\mu_1, \sigma_1^2) = (124, 256)$  a  $(\mu_2, \sigma_2^2) = (40, 1)$ . Tato volba parametrů je už z principu úplně nesmyslná, výsledek taktéž, rozložení oblastí ani zdaleka neodpovídá naší volbě.

Ať už jde o volbu konstanty  $\beta$  v apriorní hustotě nebo parametry modelu, ukázali jsme si, že jen smysluplná volba povede k nejlepším výsledkům. Nalezli jsme tedy kompromis mezi váhou apriorní informace (nedávali jsme jí takovou váhu, aby přebila informaci z reálných jasů pixelů, ale ani tak malou váhu, aby se na snímcích netvořily souvislé oblasti) a vhodnými parametry modelu jakožto bodovými odhady výběrového průměru a rozptylu odpovídajících oblastí.



## 9 Závěr

Cílem práce bylo sestavit pravděpodobnostní modely rozdělení oblastí CT snímku aorty břišní a na základě toho tento snímek segmentovat, tedy rozdělit jej na lumen (prostor, kterým teče krev), stěnu s trombem (krevní sraženinou) a vnější prostor aorty.

Abychom takový model mohli sestavit a následně provést proces segmentace, bylo třeba nejprve si závest a vysvětlit pojmy z teorie pravděpodobnosti. První kapitoly směřovaly k vysvětlení Bayesovy věty a k popsání náhodné veličiny a náhodného vektoru, včetně velice důležitého normálního rozdělení. Následně bylo popsáno markovské náhodné pole a markovský řetězec, díky němuž jsme posléze pochopili princip Markov Chain Monte Carlo metod. Kvůli správnému uchopení modelu obsahující jak absolutně spojitý, tak diskretní náhodné vektory, jsme následně udělali odbočku do teorie míry. Poslední teoretickou kapitolou pak je popsání principů bayesovské rozhodovací úlohy.

V praktické části jsme na základě výše zmíněných poznatků sestavili modely pro výplně snímku. K tomu bylo zapotřebí správně zvolit apriorní hustotu pravděpodobnosti, k čemuž nám pomohla informace o celistvosti segmentovaných oblastí. Na základě reálných hodnot jasů pixelů a apriorní hustoty pak byla určena aposteriorní hustota. Parametry modelu byly získány jako bodové odhady vhodně zvolných oblastí snímku. Pro nalezení bodového odhadu segmentace jsme díky vhodné volbě ztrátové funkce hledání minima střední hodnoty ztrátové funkce vzhledem k aposteriorní hustotě převedli na hledání maxima aposteriorní hustoty. Vzhledem k nepřívětivému tvaru této hustoty ovšem nebylo možné její maximum nalézt analyticky, využili jsme tedy Gibbsova sampleru jakožto speciálního případu Metropolisova Hastingsova algoritmu. Jedná se sice o metodu primárně určenou k numerické integraci, díky koncentrovanosti cílové hustoty však můžeme vygenerované vzorky považovat za dostatečně dobrou aproximaci segmentace. Samotný algoritmus, včetně ostatních podpůrných programů, byl naimplementován v programovacím jazyce Python.

Kromě segmentace s vhodně zvolenými parametry bylo v práci demonstrováno, jak se algoritmus chová, kdyby parametry vhodně zvoleny nebyly. Zároveň je zahrnuto i ilustrování faktu, že generované vzorky jakožto markovský řetězec opravdu dosáhnou stacionárního rozdělení.

## Reference

- [1] Martina Litschmannová. Vybrané kapitoly z pravděpodobnosti. 2011.
- [2] Radim Briš and Martina Litschmannová. Statistika I. *Pro kombinované a distanční studium*, VŠB-TU Ostrava, 2004.
- [3] Stan Z Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.
- [4] Gerhard Winkler. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*, volume 27. Springer Science & Business Media, 2012.
- [5] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [6] Jan Kracík. Poznámky k teorii míry. 2016.
- [7] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [8] Jan Kracík. Poznámky k předmětu statistika 3. 2011.
- [9] Stanislav Polzer, Jan Kracík, Tomáš Novotný, Luboš Kubíček, Robert Staffa, and Madhavan L Raghavan. Methodology for estimation of annual risk of rupture for abdominal aortic aneurysm. *Computer Methods and Programs in Biomedicine*, 2021.